

STAR 24 OCT 16 1985

The Telecommunications and Data Acquisition Progress Report 42-82

April-June 1985

E.C. Posner
Editor

(NASA-CR-176203) THE TELECOMMUNICATIONS AND
DATA ACQUISITION REPORT Progress Report,
Apr. - Jun. 1985 (Jet Propulsion Lab.)
218 p HC A10/MF A01 CSCL 17B

N86-10382
THRU
N86-10402
Unclas
29030

G3/32

August 15, 1985

NASA

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-82

April-June 1985

**E.C. Posner
Editor**

August 15, 1985



**National Aeronautics and
Space Administration**

**Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California**

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Tracking and Data Systems (OSTDS).

In geodynamics, the publication reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the search for extraterrestrial intelligence (SETI), it reports on implementation and operations for searching the microwave spectrum. The latter two programs are performed for NASA's Office of Space Science and Applications (OSSA).

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

This and each succeeding issue of the TDA Progress Report will present material in some, but not necessarily all, of the following categories:

OSTDS Tasks:

- DSN Advanced Systems
 - Tracking and Ground-Based Navigation
 - Communications, Spacecraft-Ground
 - Station Control and System Technology
 - Network Data Processing and Productivity
- DSN Systems Implementation
 - Capabilities for New Projects
 - Networks Consolidation Program
 - New Initiatives
 - Network Sustaining
- DSN Operations
 - Network Operations and Operations Support
 - Mission Interface and Support
 - TDA Program Management and Analysis
- GCF Implementation and Operations
- Data and Information Systems

OSSA Tasks:

- Search for Extraterrestrial Intelligence
- Geodynamics
 - Geodetic Instrument Development
 - Geodynamic Science

Discretionary Funded Tasks

Contents

OSTDS TASKS DSN Advanced Systems TRACKING AND GROUND-BASED NAVIGATION

A VLBI Survey at 2.29 GHz	1
D. D. Morabito, R. A. Preston, J. G. Williams, J. Faulkner, D. L. Jauncey, and G. D. Nicolson NASA Code 310-10-60-26-00	

Comparison of the Calibration of Ionospheric Delay in VLBI Data by the Methods of Dual Frequency and Faraday Rotation	11
J. A. Scheid NASA Code 310-10-60-60-00	

COMMUNICATIONS, SPACECRAFT-GROUND

A New 70-Meter Antenna Quadripod With Reduced RF Blockage	24
J. J. Cucchissi NASA Code 310-20-65-64-00	

Seismic Analysis of the Large 70-Meter Antenna, Part I: Earthquake Response Spectra Versus Full Transient Analysis	31
K. Kiedron and C. T. Chian NASA Code 310-20-65-64-00	

An Experimental $TE_{12} - TE_{11}$ Circular Waveguide Mode Converter	43
D. Hoppe NASA Code 310-20-64-22-00	

Propagation and Radiation Characteristics of a Multimode Corrugated Waveguide Feedhorn	57
D. Hoppe NASA Code 310-20-64-22-00	

Load-Deflection Tests and Computer Analyses of a High-Precision Adhesive-Bonded Antenna Reflector Panel	68
C. T. Chian and R. Levy NASA Code 310-20-65-05-10	

STATION CONTROL AND SYSTEM TECHNOLOGY

A Sequential Decoding Performance Analysis for International Comet Explorer	82
D. Divsalar NASA Code 310-30-71-83-04	

The VLSI Design of Error-Trellis Syndrome Decoding for Convolutional Codes	92
I. S. Reed, J. M. Jensen, T. K. Truong, and I. S. Hsu NASA Code 310-30-70-84-08	

High-Rate Convolutional Code Construction With the Minimum Required SNR Criterion	108
P. J. Lee	
NASA Code 310-30-71-83-02	

DSN Systems Implementation NETWORK SUSTAINING

DSN Frequency and Timing System, Mark IV-85	113
B. W. Falin	
NASA Code 314-40-41-81-13	

DSN Operations MISSION INTERFACE AND SUPPORT

JPL Emergency Support of TDRSS and Compatible Satellites	120
N. A. Fanelli	
NASA Code 314-40-22-34-01	

TDA PROGRAM MANAGEMENT AND ANALYSIS

Planning for VLA/DSN Arrayed Support to the Voyager at Neptune	125
J. W. Layland and D. W. Brown	
NASA Code 314-40-31-30-03	
A VLA Experiment – Planning for Voyager at Neptune	136
J. W. Layland, P. J. Napier, and A. R. Thompson	
NASA Code 314-40-31-30-03	
Periodic Binary Sequences With Very Good Autocorrelation Properties	143
S. Tyler and J. Loftsson	
NASA Code 055-47-11-10-01	
A Binary Sequence of Period 60 With Better Autocorrelation Properties Than the Barker Sequence of Period 13	159
J. Watkins, J. Loftsson, and S. Tyler	
NASA Code 055-47-11-10-01	

GCF Implementation and Operations

The GCF Mark IV Implementation and Beyond	163
R. A. Crowe	
NASA Code 314-20-40-00-08	

OSSA TASKS
Search for Extraterrestrial Intelligence

SETI Radio Spectrum Surveillance System 173
B. Crow, A. Lokshin, M. Marina,
and L. Ching
NASA Code 199-50-62-10-04

**Mobile VLBI and GPS Measurement of
Vertical Crustal Motion** 185
P. M. Kroger, J. M. Davidson, and
E. C. Gardner
NASA Code 692-40-40-00-07

DISCRETIONARY FUNDED TASKS

**Improved Mapping of Radio Sources From
VLBI Data by Least-Squares Fit** 199
E. R. Rodemich
NASA Code 404-00-73-02-55

Referees 211

N 86 - 1 0 3 8 3

A VLBI Survey at 2.29 GHz

D. D. Morabito, R. A. Preston, J. G. Williams, and J. Faulkner
Tracking Systems and Applications Section

D. L. Jauncey
CSIRO Division of Radio Physics, Sydney, Australia

G. D. Nicolson
CSIR, Johannesburg, South Africa

The Deep Space Network is establishing a high-accuracy VLBI celestial reference frame. This article presents the results of a search for suitable radio sources to be used in constructing this frame. The VLBI observations using DSN baselines at 2.29 GHz with fringe spacings of about 3 milliarcseconds have been performed on 1398 radio sources spread over the entire sky. Of those, 917 sources were detected including 93% of the identified BL Lacertae objects, 86% of Quasars and 36% of galaxies. The resulting catalog of compact radio sources is also useful for various astrophysical studies and in the formation of VLBI celestial reference frames.

I. Introduction

This article presents the results of a systematic VLBI full-sky survey undertaken to establish a comprehensive catalog of ultracompact celestial radio sources. The survey was conducted by performing 2.29 GHz VLBI observations on known radio sources to search for compact structure. Of 1398 radio sources observed, 917 sources have been detected. Arcsecond positions for 787 of the detected sources have been previously determined from the VLBI survey data (Refs. 1, 2, 3) and are being used to identify optical counterparts (Refs. 4, 5, 6).

The results of this survey are presently being utilized to form a VLBI reference frame of 100-200 sources by determining precise relative positions ($0''.01-0''.01$) (see Ref. 7). Such celestial reference frames (see also Ref. 8) will be at least an order of magnitude more precise than previous stellar frames and are nearly inertial since the extragalactic sources are

without measurable proper motions. They enable significant advances in various geodetic and astrometric studies (e.g., crustal plate dynamics, earth rotational irregularities, planetary dynamics, interplanetary spacecraft navigation). A similar, but deeper, VLBI survey of the ecliptic zone has been previously published (Ref. 9).

The VLBI survey is also useful for studying the characteristics of compact radio sources. The detected survey sources coincide with the cores of quasars and galaxies. Understanding the nature of these energetic cores is crucial in unraveling the origin and evolution of the objects in which they reside. The catalog will not only serve as a reference list for observers, but it can be used in statistical studies of radio source properties and cosmological theories. Toward these ends, the catalog has been supplemented with optical identifications, optical magnitudes, redshifts and radio spectral indices derived from the literature.

II. Sample Selection and Completeness

Candidate sources were selected primarily from the Parkes survey (Ref. 10) and the NRAO-Bonn survey (Ref. 11) which together span the entire sky ($|b^{\text{II}}| > 10$ degrees). These surveys both provide total flux density measurements at 2.7 and 5.0 GHz for most sources. The sample observed with VLBI covers the full sky and was chosen largely on the basis of criteria placed on total flux density $S(S_{2.7})$ and spectral index $\alpha(\alpha_{2.7}^{5.0}; S \equiv S_o f^\alpha)$, neglecting temporal variability. For example, for those sources for which the Parkes and NRAO-Bonn surveys give total flux densities at both 2.7 and 5.0 GHz, 100% of the sources were observed for which $S \geq 1.0$ Jy and $\alpha \geq 0.0$ (114 sources observed, 105 detected), and 89% for which $S \geq 0.5$ Jy and $\alpha \geq -0.5$ (717 of 805 sources observed, 592 detected).

Also observed were 681 weaker or steeper spectrum sources from the Parkes and NRAO-Bonn surveys, as well as from the general literature. Our sample was intended to be purely extragalactic, and identified galactic sources were eliminated from the sample. However, some of the optically unidentified sources that met our sample criteria could be galactic. Such sources are highly unlikely to have been detected with VLBI at our angular resolution and sensitivity.

Completeness of the observed sample is difficult to estimate, not only due to temporal variability but also because the two finding surveys had different levels of completeness for different sky regions, lacked two frequency information for all sources, and had different primary survey frequencies. Neglecting temporal variability, both finding surveys are nearly complete for $S \geq 1.0$ Jy and $\alpha \geq -0.5$, resulting in a combined completeness of more than 97% for the sky area covered. The spectral index criterion is necessary because the NRAO-Bonn survey frequency was 5.0 GHz, not 2.7 GHz. Based on these sample criteria, the VLBI survey is estimated to be 93% complete, again neglecting temporal variability, with a total of 312 sources observed. Because the flux density limits of the finding surveys varied depending on sky region, estimates of completeness for sources with lower total flux densities do not apply to the entire sky (see Table 1).

III. The Observations

The observations were performed at 2.29 GHz with pairs of antennas on California-Spain, California-Australia, and Australia-South Africa baselines (see Table 2) during 68 different observing sessions between 1974 and 1983 (see Table 3). Right circular polarization was received and data were recorded on the NRAO Mark II system (Ref. 12).

The fringe spacing sampled ranged from 2.5 to 4.1 milliarcseconds. For the mean fringe spacing of 3.3 milliarcseconds, the normalized fringe visibility of a Gaussian source varies from 0.9 to 0.1 as the half-intensity diameter increases from 0.5 to 2.2 milliarcseconds.

The 5σ detection limit in correlated flux density was generally ~ 0.1 Jy. The corresponding random uncertainty in detected source strength is ~ 0.02 Jy, but systematic errors at about the 10% level dominate the random contribution for most sources. To ensure that few compact radio components would be missed due to a priori source position errors, the sky was completely searched within 0.5 arcminutes of all nominal source positions by cross-correlating over an appropriate range of delay and delay rate.

Total flux densities at 2.29 GHz were also measured for most sources at the time of VLBI measurement by means of on-off measurements with a noise-adding radiometer. The random uncertainties in total flux density measurements typically range from 0.03 to 0.3 Jy, with systematic errors in antenna sensitivity being $\sim 3\%$.

IV. Results

Of 1398 sources observed, 917 (or 66%) were detected with VLBI; 83% of the observed sources with $S \geq 0.5$ Jy and $\alpha \geq -0.5$ were detected. Figure 1 is an equal area sky distribution plot of the detected objects. Sparsity near the galactic plane is evident. Figure 2 is a correlated flux density histogram of the detected objects. There are 49 sources with correlated flux densities greater than 1 Jansky, and 227 sources with correlated flux densities greater than 0.5 Jansky.

Detection statistics as a function of optical identification type appear in Table 4. Detection statistics as a function of general optical class appear in Table 5: 93% of identified BL Lacertae objects were detected, 86% of QSOs were detected and 36% of galaxies were detected.

Figure 3 displays a sample page of the survey results. Figure 4 displays a sample page of the reference table to the supplementary data given in Fig. 3. Figures 3 and 4 will appear in their entirety in a separate report. Descriptions of the entries to Fig. 3 appear below.

<u>Column</u>	<u>Description</u>
1	Source name
2/3	1950.0 positions. Asterisked positions are determined from the VLBI survey data and have typical uncertainties of 1 arcsecond (see Refs. 1-3). Other

<u>Column</u>	<u>Description</u>
	positions are from the literature, and in most cases, errors are < 30 arcseconds.
4	Spectral indices between 2700 to 5000 MHz followed by corresponding reference number (see Fig. 4). A few existing compilations of redshifts, optical identifications and optical magnitudes were useful aids in preparing our catalog (Refs. 63, 84, 86, 111 and 232 in Fig. 4). However, in almost all cases we have drawn values for these quantities from original references to enhance accuracy. A star following the reference number indicates a questionable or conflicting value, and is explained in the notes to Fig. 3. For many Southern Hemisphere sources, the optical characteristics were obtained from an optical identification program which utilized the radio source positions determined by our survey (Refs. 4, 5, 6).
5	Redshifts followed by corresponding reference number (see Fig. 4).

<u>Column</u>	<u>Description</u>
6	Optical identifications followed by corresponding reference number (see Fig. 4). Optical identification codes are defined in Table 4.
7	Optical magnitudes followed by the corresponding reference number (see Fig. 4). These values may be visual, blue or red.
8	Experiment codes as defined in Table 3.
9	Measured 2.29 GHz total flux density (Jansky).
10	Measured 2.29 GHz correlated flux density (Jansky). Values for 17 ecliptic sources marked by asterisks are from Ref. 9.
11	Visibility is defined as the correlated flux density divided by the total flux density.
12	East-west (u) and north-south (v) spatial frequencies of the observations in units of 10^6 wavelengths.

Computer readable versions of the catalog are available upon request.

Acknowledgments

We are indebted to the instrumentation and observing assistance supplied by L. J. Skjerve, D. J. Spitzmesser, and the operations personnel of the Deep Space Network, the Australian National Radio Observatory at Parkes, Australia, and the Hartebeesthoek Radio Astronomy Observatory, South Africa. We are also grateful to A. W. Harris, E. Padua, M. A. Slade, G. B. Trager and J. Kidder for computational aid.

References

1. Morabito, D. D., Preston, R. A., Slade, M. A., and Jauncey, D. L. 1982, "Arcsecond Positions for Milliarcsecond VLBI Nuclei of Extragalactic Radio Sources. I. 546 Sources," *A.J.*, 87, 517-527.
2. Morabito, D. D., Preston, R. A., Slade, M. A., Jauncey, D. L., and Nicolson, G. D. 1983, "Arcsecond Positions for Milliarcsecond VLBI Nuclei of Extragalactic Radio Sources. II. 207 Sources," *A.J.*, 88, 1138-1145.
3. Morabito, D. D., Wehrle, A. E., Preston, R. A., Linfield, R. P., Slade, M. A., Faulkner, J., and Jauncey, D. L. "Arcsecond Positions for Milliarcsecond VLBI Nuclei of Extragalactic Radio Sources: III. 74 Sources," April 1985, *A.J.*, 90, 590-594.
4. Jauncey, D. L., Batty, M. J., Wright, A. E., Peterson, B. A., and Savage, A. 1984, "Redshifts of Southern Radio Sources. VI," *Ap. J.*, 286, 498-502.
5. Jauncey, D. L., Savage, A., Morabito, D. D., and Preston, R. A. 1985, "Optical Identifications of Compact Radio Sources: Candidates for the Construction of an Extragalactic Radio/Optical Reference Frame," in preparation.
6. Savage, A., Jauncey, D. L., Batty, M. J., Gulkis, S., Morabito, D. D., and Preston, R. A. 1983, "Optical Identifications of Radio Sources with Accurate Positions Using the UKST IIIa-J Plates," *Proc. IAU Colloquium No. 78, Astronomy with Schmidt Type Telescopes*, Asiago, Italy 481-487.
7. Fanselow, J. L., Sovers, O. J., Thomas, J. B., Purcell, G. H., Cohen, E. J., Rogstad, D. H., Skjerve, L. J., and Spitzmesser, D. J. 1984, "Radio Interferometric Determination of Source Positions Utilizing Deep Space Network Antennas - 1971 to 1980," *A. J.* 89, 987-998.
8. Ma, C., Clark, T. A., and Shaffer, D. B. 1981, "Mark III VLBI: Astrometry and Epoch J2000.0," *Bull. AAS*, 13, 899-899.
9. Wehrle, A. E., Morabito, D. D., and Preston, R. A. 1984, "Very Long Baseline Interferometry Observations of 257 Extragalactic Radio Sources in the Ecliptic Region," *A. J.* 89, 336-341.
10. Bolton, J. G. 1979, Magnetic Tape Version of Unpublished Master Parkes Catalog, maintained by CSIRO, Division of Radio Physics, Sydney, Australia.
11. Kuhr, H., Nauber, U., Pauliny-Toth, I., and Witzel, A. 1979, "A Catalogue of Radio Sources," Max-Planck-Institut fur Radioastronomie, *Preprint No. 55*, 1-337.
12. Clark, B. G. 1973, "The NRAO Tape Recorder Interferometer System," *Proc. IEEE*, 61, 1242-1248.

Table 1. Completeness estimates for various sample criteria

Declination Range, deg	Flux Density Limit ($S_{2.7}$), Jy	Spectral Index Limit (α 5.0 / 2.7)	Number of Sources Observed*	Number of Sources Detected*	Completeness of Sample, %
-90 to +90	> 1.0	> -0.5	312	258	93
-90 to +27	> 0.65	> -0.5	396	336	85
+70 to +90	> 0.5	> -0.5	30	29	97

*Includes small adjustment to account for estimated number of sources with total flux density given at only one frequency which would have $\alpha > -0.5$.

Table 2. Participating observatories

Location	Designation	Diameter, m	Baseline Length	
			Kilometers $\times 10^3$	Wavelengths $\times 10^6$
Hartebeesthoek, S. Africa	HT	26	9.7	75
Parkes, Australia	PK	64		
Tidbinbilla, Australia	43	64		
	42	26	10.6	81
Goldstone, California	14	64		
	13	26	8.4	64
Madrid, Spain	63	64		
	61	26		

Table 3. Experiment summary

Experiment Code	Date			Observatories	Experiment Code	Date			Observatories
	YR	MN	DY			YR	MN	DY	
1	74	07	31	14 42	35	78	03	05	13 63
2	74	08	13	14 42	36	78	03	14/15	13 63
3	75	06	17	14 42	37	78	03	21	13 63
4	75	06	18	14 42	38	78	04	04	13 63
5	75	08	23/24	14 42	39	78	04	16	13 43
6	75	08	23	14 62	40	78	05	30/31	13 43
7	75	09	15	14 42	41	78	06	29/30	13 43
8	75	09	21	13 63	42	78	10	18	13 43
9	75	10	26	13 43	43	80	2	1	14 43
10	76	05	27	14 63	44	80	2	27/28	14 43
11	76	11	09/10	13 63	45	80	3	2	13 43
12	76	11	11	13 63	46	80	3	3	13 63
13	76	11	14	13 43	47	80	3	12/13	13 63
14	77	01	28	13 43	48	80	3	14/15	13 63
15	77	02	12	13 43	49	80	3	19	13 43
16	77	02	20	13 43	50	80	3	26	13 43
17	77	02	21	13 43	51	80	3	27	13 63
18	77	02	23	13 43	52	80	4	24-27	PK HT
19	77	02	25	13 43	53	80	4	26	43 HT
20	77	04	21	13 43	54	80	6	19	14 43
21	77	04	22	13 43	55	81	1	21	13 43
22	77	06	15	13 43	56	81	1	25	13 43
23	77	09	11	14 61	57	81	1	31	13 43
24	77	09	28/29	13 63	58	81	3	1	14 43
25	77	10	11	13 63	59	81	4	22	14 43
26	77	10	27/28	13 43	60	81	5	8	14 43
27	77	11	01	13 43	61	81	10	23	13 43
28	77	11	21	13 43	62	81	10	26	13 43
29	77	12	02	13 43	63	81	11	1	13 43
30	77	12	11	13 43	64	82	2	14	43 HT
31	77	12	12	13 43	65	82	2	17	43 HT
32	77	12	13/14	13 43	66	82	2	19/20	43 HT
33	78	01	09/10	13 43	67	82	4	20	43 HT
34	78	02	20/21	13 43	68	83	6	21	13 43

Table 4. Optical identification codes with VLBI detection statistics

Optical ID	Description	Number of Detections	Number of Nondetections	Percent Detected
B	Blue Stellar Object	34	6	85
CG	Compact Galaxy	1	2	33
D	Diffuse Galaxy	6	9	40
DB	DB Galaxy	3	8	27
D4	D4 Galaxy	0	2	0
E	Elliptical Galaxy	10	28	26
E0	E0 Galaxy	3	4	43
E1	E1 Galaxy	1	2	33
E2	E2 Galaxy	2	2	50
E3	E3 Galaxy	1	1	50
E4	E4 Galaxy	3	2	60
E5	E5 Galaxy	0	1	0
EF	Empty Field	84	65	56
G	Galaxy	48	97	33
L	BL Lacertae Object	56	4	93
N	Neutral Stellar Object	14	3	82
NG	N-Galaxy	11	5	69
N2	Type 2 N-Galaxy	0	2	0
PG	Probable Galaxy	10	9	53
PQ	Probable QSO	41	25	62
Q	Quasi-Stellar Object	503	67	88
R	Red Stellar Object	7	1	88
S	Spiral Galaxy	0	2	0
SB	Sb Galaxy	0	1	0
SC	Sc Galaxy	0	1	0
SG	Seyfert Galaxy	2	1	67
S0	S0 Galaxy	1	1	50
U	Unidentified Object	2	3	40
(Blank)	No Information	74	127	37
Totals		917	481	66

Table 5. Optical class VLBI detection statistics

Class	Number of Detections	Number of Nondetections	Percent Detected
Galaxies and Probable Galaxies	102	180	36
QSOs and Probable QSOs	544	92	86
BL Lacertae	56	4	93
Stellar Objects	55	10	85
Empty Fields	84	65	56
No Information and Unidentified	76	130	37
Totals	917	481	66

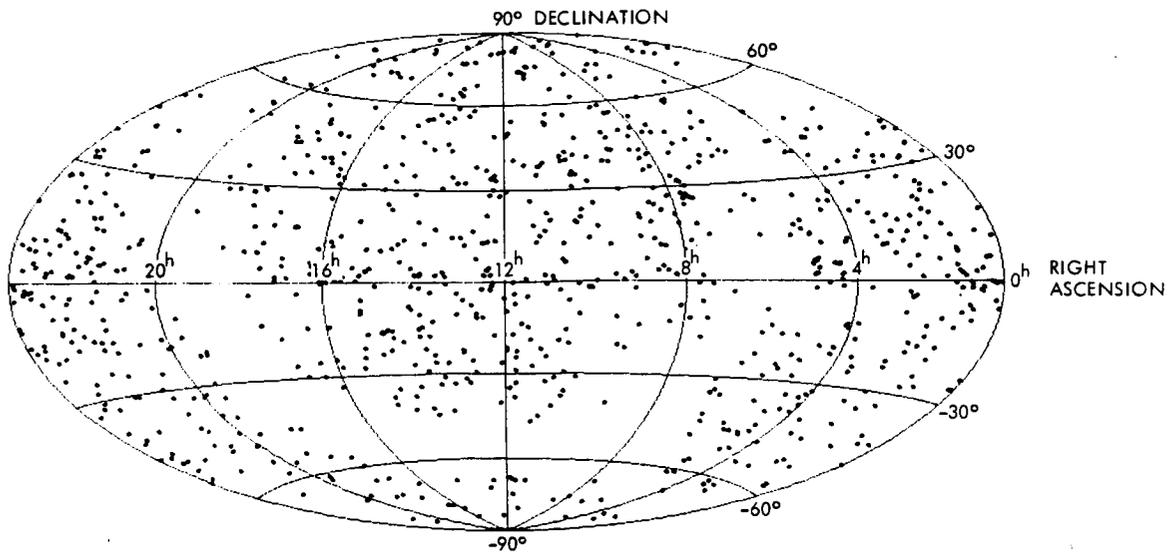


Fig. 1. Sky distribution of sources detected with VLBI

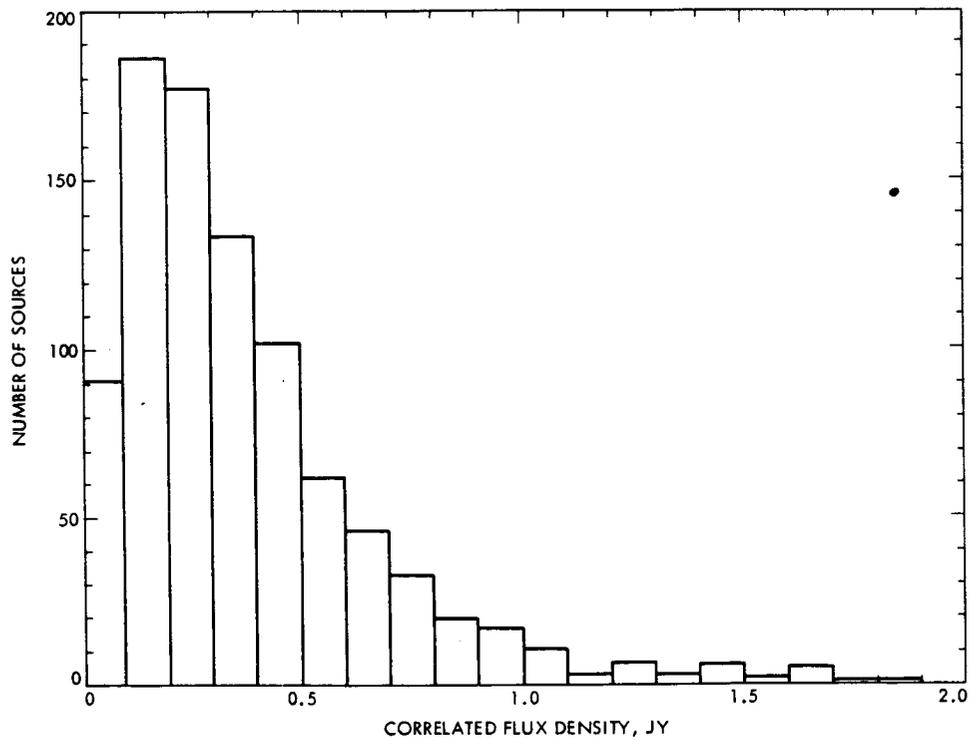


Fig. 2. Histogram of correlated flux densities for detected sources

(1) SOURCE NAME	(2) RIGHT ASCENSION HR MIN SEC	(3) DECLINATION DEG MN SEC	(4) SPECTRAL INDEX	(5) RED SHIFT	(6) OPTICAL ID	(7) OPTICAL MAG	(8) EXPT CODE	(9) TOTAL FLUX DENSITY (JY)	(10) CORRELATED FLUX DENSITY (JY)	(11) VISIBILITY	(12) U V (10**6 MWLNS)
P 0000-550	0 0 35.7	-55 1 58	-0.5 96		E 96	14.5 96 52		<0.12			-68.4 -19.2
P 0002-478	0 2 2.9	-47 53 2 *	-0.2 96		0 174	19.0 199 52		0.47 +/-0.04			-69.9 -13.6
GC 0003+38	0 3 22.4	+38 3 33 *	-0.3 44		0 44	19.4 13 28	0.6 +/-0.3		0.6 +/-0.3		-49.9 -62.8
P 0003-56	0 3 26.6	-56 43 10	-0.8 96		0 96	19.5 96 52		<0.12			-71.3 18.1
P 0003-42	0 3 28.0	-42 52 18	-1.0 102	1.85 107	0 103	19.5 103 52		<0.12			-71.9 14.5
NRAD 5	0 3 40.2	-6 40 17 *	0.0 63	0.347 104*	PG 104	18.5 104 7	1.69+/-0.07	0.46 +/-0.05	0.27 +/-0.03		-58.2 -56.2
3C 2	0 3 48.8	-0 21 6	-0.8 63	1.037 5	0 5	19.5 5 54	2.63+/-0.07	0.062+/-0.006	0.023+/-0.002		-54.1 -56.0
P 0003-83	0 3 54.6	-83 22 22	-0.5 102		0 102	19.0 102 52		<0.12			-65.9 33.3
P 0005-239	0 5 27.5	-23 56 0 *	-0.1 39	1.407 39	0 16	17.0 30 33		0.43 +/-0.05			-58.2 -53.0
P 0005-262	0 5 53.5	-26 15 53 *	0.0 16		0 16*	20. 16 34		0.21 +/-0.03			-51.6 -62.3
III ZW 2	0 7 57.9	+10 41 30 *		0.089 86	CG 202	15.4 202 33		0.15 +/-0.03			-57.0 -52.7
GC 0007+17	0 7 59.4	+17 7 38 *	0.4 102	1.201 30	0 30	18.0 30 11		0.6 +/-0.1			63.5 0.8
P 0008-42	0 8 21.8	-42 9 47	-1.3 63		PG 160	22.0 160 32	3.1 +/-0.3	<0.15	< 0.048		-41.6 -69.0
P 0008-264	0 8 28.9	-26 29 15 *	+0.3 16	1.096 104	PG 16	19. 16 34	0.58+/-0.06	0.53 +/-0.05	0.9 +/-0.1		-50.3 -63.3
P 0010+00	0 10 37.1	+0 35 3	-1.0 102		EF 102		1.08+/-0.04	<0.02	< 0.018		-54.0 -55.9
GC 0010+40	0 10 54.3	+40 34 57 *	-0.3 63		0 115	17.9 115 38		0.34 +/-0.04			43.3 -28.2
P 0011-046	0 11 20.6	-4 40 33 *	+1.1 79		PG 79	19.5 79 33		0.28 +/-0.03			-58.3 -55.8
GC 0012+31	0 12 29.9	+31 59 33 *	-0.5 63		PG 80	19. 80 34	0.29+/-0.05	0.22 +/-0.03	0.8 +/-0.2		-40.5 -69.8
P 0013-00	0 13 37.4	-0 31 53 *	-0.4 102		R 13	19.8 13 61		0.33 +/-0.02*			-57.0 -56.0
0014+81	0 14 4.5	+81 18 29 *	-0.2 63		0 63	16.5 63 51	1.0 +/-0.3	0.35 +/-0.03	0.4 +/-0.1		-57.9 27.5
0016+73	0 16 54.1	+73 10 52	0.2 63		0 63	18.0 63 51	1.5 +/-0.1	0.54 +/-0.04	0.36 +/-0.04		-49.2 40.3
0018+72	0 18 34.5	+72 56 4 *	-0.7 63		0 21	21.1 21 44	0.8 +/-0.2	0.11 +/-0.01	0.13 +/-0.04		-50.0 39.3
P 0019-00	0 19 51.7	-0 1 42 *	-0.9 63		L 186	19.2 186 33	2.24+/-0.06	0.059+/-0.004	0.026+/-0.002		-55.9 -56.1
P 0019+058	0 19 58.3	+5 51 22 *	+0.2 75		EF 174			0.52 +/-0.06			-57.4 -54.6
P 0022-423	0 22 15.4	-42 18 41 *	-0.8 102					0.75 +/-0.04			-69.9 19.1
08 337.7	0 22 46.7	+39 2 59 *	0.2 44		0 44	19.8 13 24	0.7 +/-0.2	0.45 +/-0.07	0.6 +/-0.2		54.1 24.4
P 0022-60	0 22 54.4	-60 45 6	-1.1 96		0 160	20.5 160 32	6.4 +/-0.3	<0.05	< 0.022		-70.4 19.8
P 0023-26	0 23 17.9	-26 18 45	-0.8 63		0 45	19. 8 28	0.9 +/-0.2	0.37 +/-0.04	0.4 +/-0.1		-51.8 -61.3
08 338	0 24 2.8	+34 52 6 *	0.1 63		PG 116	19.0 96 52		<0.12			-66.4 -20.7
P 0024-495	0 24 16.3	-49 35 21	-0.6 96		0 147	20.2 147 24	1.8 +/-0.2	0.16 +/-0.06	0.09 +/-0.03		57.7 18.8
08 343	0 26 34.8	+34 39 58 *	-0.3 44		EF 63	19.5 75 33		0.40 +/-0.05	0.26 +/-0.04		-56.3 -54.2
P 0027+056	0 27 11.4	+5 38 5 *	+0.4 75		0 5	19.4 5 44	0.8 +/-0.1	0.21 +/-0.02	< 0.026		-49.8 39.1
0027+70	0 27 17.0	+70 21 6 *	0.0 63		D 102	19. 102 68		<0.02			-56.6 -56.4
P 0028-01	0 28 58.5	-1 17 22	-0.8 102		E 5	19.6 5 54	0.66+/-0.03	0.22 +/-0.01	0.052+/-0.003		-53.4 -57.0
P 0030+19	0 30 1.2	+19 37 12	0.8 102		E 5	19.6 5 54	4.28+/-0.11	0.33 +/-0.01	< 0.217		-49.2 -63.6
P 0034-01	0 34 30.6	-1 25 38 *	-0.8 102	0.073 21	E 5	17.6 21 44	2.90+/-0.08	0.029+/-0.002	0.010+/-0.001		-57.4 -56.4
P 0035+23	0 35 19.8	+23 50 42 *	-0.7 102	2.27 36	PG 200	19. 200 33		0.10 +/-0.02			-56.0 -57.8
GC 0035+12	0 35 41.9	+12 11 2 *	-0.5 63		B 32	17.0 32 33		0.14 +/-0.02			-56.6 -57.8
P 0035-02	0 35 47.2	-2 24 9 *	-0.7 102	0.220 176	E 5	19.6 5 54	4.28+/-0.11	0.33 +/-0.01	0.052+/-0.003		-53.4 -57.0
P 0035-216	0 36 0.4	-21 36 34	0.2 16		0 16	19. 16 32	0.6 +/-0.2	<0.12	< 0.217		-49.2 -63.6
P 0035-39	0 36 2.3	-39 16 13	-1.1 102	0.592 25	0 71	16.5 71 52		<0.12			-71.5 15.3
P 0036-42	0 36 30.0	-62 48 12	-0.9 96		E 96	18.5 96 52		<0.12			-70.7 20.6
P 0036+03	0 36 44.2	+3 3 25 *	-1.0 102	0.014 177	E 2 5	13.5 5 54	1.38+/-0.05	0.026+/-0.002	0.019+/-0.002		-54.9 -59.2
P 0038-326	0 38 5.1	-32 41 40	0.6 73		0 32	17.0 32 33		<0.14			-41.7 -69.2
NB 89. 01	0 38 20.0	+89 12 36			B 1.3	1.3 +/-0.2		<0.11	< 0.085		-63.8 -5.8

Fig. 3. Sample page of VLBI survey results

ORIGINAL PAGE IS
OF POOR QUALITY

1. Burbidge, E. M. (1967). *Astrophys. J. Letters*, 149, L51.
2. Bolton, J. G., Kinman, T. D., and Wall, J. V. (1968). *Astrophys. J. Letters* 154, L105.
3. Searle, L., and Bolton, J. G. (1968). *Astrophys. J. Letters* 154, L101.
4. Veron, M. P. (1971). *Astron. Astrophys.* 11, 1.
5. Wall, J. V., Shimmins, A. J., and Merheli, J. K. (1971). *Aust. J. Phys. Suppl.* 19, 1.
6. Fanaroff, B. L., and Blake, G. M. (1972). *Mon. Not. R. astr. Soc.* 157, 41.
7. Baldwin, J. A., Burbidge, E. M., Hazard, C., Murdoch, H. S., Robinson, L. B., and Wampler, E. J. (1973). *Astrophys. J.* 185, 739.
8. Browne, I. W. A., Crowther, J. H., and Adgie, R. L. (1973). *Nature* 244, 146.
9. Peterson, B. A., Bolton, J. G., and Shimmins, A. J. (1973). *Astrophys. Letters* 15, L109.
10. Sandage, A. (1973). *Astrophys. J.* 180, 697.
11. Wills, B. J., Wills, D., and Douglas, J. N. (1973). *Astron. J.* 78, 521.
12. Hoskins, D., Murdoch, H. S., Adgie, R. L., Crowther, J. H., and Gent, H. (1974). *Mon. Not. R. astr. Soc.* 166, 235.
13. Johnson, K. H. (1974). *Astron. J.* 79, 1006.
14. Strittmatter, P. A., Carswell, R. F., Gilbert, G., and Burbidge, E. M. (1974). *Astrophys. J.* 190, 509.
15. Wills, D., and Wills, B. J. (1974). *Astrophys. J.* 190, 271.
16. Bolton, J. G., Shimmins, A. J., and Wall, J. V. (1975). *Aust. J. Phys. Suppl.* 34, 1.
17. Colla, G., Fanti, C., Fanti, R., Gioia, I., Lari, C., Lequeux, J., Lucas, R., and Ulrich, M. H. (1975). *Astron. Astrophys. Suppl.* 20, 1.
18. Condon, J. J., Balonek, T. J., and Jauncey, D. L. (1975). *Astron. J.* 80, 887.
19. Edwards, T., Kronberg, P. P., and Menard, G. (1975). *Astron. J.* 80, 1005.
20. Fanti, C., Fanti, R., Ficarra, A., Formiggini, L., Giovannini, G., Lari, C., and Padrielli, L. (1975). *Astron. Astrophys. Suppl.* 19, 143.
21. McEwan, N. J., Browne, I. W. A., Crowther, F. H. (1975). *Mem. R. astr. Soc.* 80, 1.
22. Condon, J. J., Balonek, T. J., and Jauncey, D. L. (1976). *Astron. J.* 81, 913.
23. Kuhr, H. (1977). *Astron. Astrophys. Suppl.* 29, 139.
24. Peterson, B. A., Bolton, J. G., and Savage, A. (1976). *Astrophys. J. Letters* 17, L137.
25. Peterson, B. A., Jauncey, D. L., Wright, A. E., and Condon, J. J. (1976). *Astrophys. J.* 207, L5.
26. Savage, A., Browne, I. W. A., and Bolton, J. G. (1976). *Mon. Not. R. astr. Soc.* 177, 77P.
27. Smith, H. E., Spinrad, H., and Smith, E. D. (1976). *Publ. Astron. Soc. Pacific* 88, 621.
28. Veron, M. P., Veron, P., Adgie, R. L., and Gent, H. (1976). *Astron. Astrophys.* 47, 401.
29. Wills, B. J. (1976). *Astron. J.* 81, 1031.
30. Wills, D., and Wills, B. J. (1976). *Astrophys. J. Suppl.* 31, 143.
31. Boksenberg, A., Briggs, S. A., Carswell, R. F., Schmidt, M., and Walsh, D. (1976). *Mon. Not. R. astr. Soc.* 177, 43P.
32. Condon, J. J., Hick, P. D., and Jauncey, D. L. (1977). *Astron. J.* 82, 692.
33. Merheli, J. K., Shimmins, A. J., and Bolton, J. G. (1968). *Aust. J. Phys.* 21, 523.
34. Ghigo, F. D. (1977). *Astrophys. J. Suppl.* 35, 359.
35. Schmidt, M. (1977). *Astrophys. J.* 217, 358.
36. Smith-Haenni, A. L. (1977). *Astron. Astrophys. Suppl.* 27, 205.
37. Spinrad, H., Westphal, J., Kristian, J., and Sandage, A. (1977). *Astrophys. J. Letters* 216, L87.

Fig. 4. Sample page of references to spectral indices, red shifts, optical identifications and optical magnitudes

N86-10384

Comparison of the Calibration of Ionospheric Delay in VLBI Data by the Methods of Dual Frequency and Faraday Rotation

J. A. Scheid

Tracking Systems and Applications Section

When both S-band and X-band data are recorded for a signal which has passed through the ionosphere, it is possible to calculate the ionospheric contribution to signal delay. In Very Long Baseline Interferometry (VLBI) this method is used to calibrate the ionosphere. In the absence of dual frequency data, the ionospheric content measured by Faraday rotation, using a signal from a geostationary satellite, is mapped to the VLBI observing direction. The purpose of this article is to compare the ionospheric delay obtained by these two methods. The principal conclusions are: 1) the correlation between delays obtained by these two methods is weak; 2) in mapping Faraday rotation measurements to the VLBI observing direction, a simple mapping algorithm which accounts only for changes in hour angle and elevation angle is better than a more elaborate algorithm which includes solar and geomagnetic effects; 3) fluctuations in the difference in total electron content as seen by two antennas defining a baseline limit the application of Faraday rotation data to VLBI.

I. Introduction

In VLBI, two or more antennas track the same extragalactic source simultaneously, and at each station the received signal is digitized and recorded on magnetic tape. The tapes from a pair of stations are then correlated to determine the time delay between the arrival of a wave front at one station and the arrival of the same wave front at the other. The total delay has geometric, tropospheric, ionospheric and instrumental components,

$$\tau = \tau_G + \tau_T + \tau_{Ib} + \tau_{Cb}, \quad b = S, X \quad (1)$$

The geometric delay τ_G is nondispersive. For the purposes of this analysis, tropospheric delay τ_T is considered nondispersive.

Ionospheric delay τ_I is dispersive. Instrumental delay τ_C includes the station clock offsets as well as dispersive components due to differences in the antennas and electronics at the two stations. Accordingly, the ionospheric and instrumental delays carry the subscript b to indicate the frequency band of the observation.

The excess phase delay due to the propagation of radiation through a dispersive medium is given for a single frequency by

$$\tau = \frac{1}{c} \int dS (n - 1) \quad (2)$$

where n is the index of refraction, c is the velocity of light in vacuum and the integration is over the ray path (see Ref. 1 for

basic ionospheric formulas). The major effect of the ionosphere is due to a plasma of free electrons. Under the assumption that damping due to electron collisions is negligible, the index of refraction at a point in the ionosphere is given by

$$n = (1 - r)^{1/2} \quad (3)$$

$$r = \frac{f_p^2}{f_b^2 + f_b f_H \cos \phi}$$

where f_b is the observing frequency in band b , f_p is the plasma frequency, f_H is the frequency of precession of an electron in the geomagnetic field, and ϕ is the angle between the wave normal and the direction of the magnetic field. Plasma frequency in Hertz is related to electron density d in electrons per cubic meter by

$$f_p = 8.984 d^{1/2} \quad (4)$$

Expanding the expression for n , the quantity $n - 1$ to be integrated along the ray path takes the form

$$n - 1 = -\frac{1}{2} r - \frac{1}{8} r^2 \dots \quad (5)$$

Normally for the plasma of the ionosphere, $f_H < f_p < 15$ MHz. For frequencies of interest in VLBI experiments, $f_p \ll f_b$. Consequently, a first order approximation to $n - 1$ is obtained by neglecting the geomagnetic field and retaining only the terms quadratic in r ,

$$n - 1 = -\frac{1}{2} \frac{f_p^2}{f_b^2} \quad (6)$$

In this approximation, the group delay at a single station caused by the ionosphere is given by

$$\frac{40.31}{c f_b^2} \int dh \frac{D(h)}{\cos \alpha} \quad (7)$$

where h is elevation in meters, $D(h)$ is the electron density profile in electrons per cubic meters, and α is the angle between the local vertical and the tangent to the ray to the source. Neglecting the bending of the ray, Eq. (7) is approximated by

$$\frac{40.31}{c f_b^2} \frac{TEC}{\cos \alpha} \quad (8)$$

where TEC is the columnar total electron content at zenith for station i .

For VLBI observations, the quantity of interest is the difference between the delays at the two stations with respect to a common reference point,

$$\tau_{Ib} = \frac{40.31}{c f_b^2} \left[\frac{TEC_1}{\cos \alpha_1} - \frac{TEC_2}{\cos \alpha_2} \right] \quad (9)$$

II. Dual Frequency Method

In reference to Eq. (9), the total observed delay may be approximated by

$$\tau_b = \tau_o + \frac{N}{f_b^2} \quad (10)$$

Thus a measurement of the delay at two frequencies makes possible the determination of the constants τ_o and N so that τ_b may be removed from the total delay at both frequencies.

In processing VLBI data, the ionospheric contribution to delay is determined in the program CALIBRATE for both the dual frequency and Faraday rotation methods. The frequencies used in CALIBRATE are weighted averages of the channel frequencies in each band. In the dual frequency approach, the quantities

$$A = \frac{(f_S^2 \tau_S - f_X^2 \tau_X)}{(f_S^2 - f_X^2)} \quad (11)$$

$$B = f_S^2 f_X^2 \frac{(\tau_X - \tau_S)}{(f_S^2 - f_X^2)}$$

are computed (Ref. 2). If Eq. (10) were exact, the quantities A and B would equal τ_o and N respectively. Contributions to dispersion other than those of Eq. (10), however, introduce frequency dependent terms into A and B . Such contributions are present but are known or assumed to be constant in time. Terms known to be small were omitted by truncating the expansion of the index of refraction and neglecting the geomagnetic field.

The data processed by CALIBRATE have been corrected by phase calibration to remove instrumental effects. The component of instrumental phase shift which is independent of frequency appears in the delay observable as a term proportional to $1/f$. If this is not eliminated by phase calibration, it will contaminate the determination of both τ_o and N . Any instrumental delay inversely proportional to frequency squared which is not removed by phase calibration contributes directly to the value of N determined from the data.

III. Faraday Rotation Method

Faraday rotation refers to the rotation of the axis of the polarization ellipse of an electromagnetic wave as it propagates through a magnetized plasma. The total Faraday rotation in radians due to passage of the wave through the ionosphere in the presence of the geomagnetic field is given by

$$\frac{\pi}{cf^2} \int dh f_H f_P^2 \frac{\cos \phi}{\cos \alpha} \quad (12)$$

Faraday rotation is measured along the ray path of a signal from a geostationary Applications Technology Satellite (ATS) to a ground receiving station at Goldstone. Total electron content for this slanted ray path is then computed and mapped to the zenith at the zenith reference point. The zenith reference point is defined to be the point along the ray path between the ATS and the ground station at a reference altitude, typically 350 km. The result, the reference point zenith electron content, is the form of the Faraday rotation data which is input to CALIBRATE.

In CALIBRATE these data are mapped to the ray path through the ionosphere along the lines of sight from the VLBI stations to the source. This mapping is designed to account for several differences between the conditions of the Faraday rotation measurement and those of the VLBI observation (Ref. 3). These include factors to account for changes in hour angle, solar-zenith angle, geomagnetic latitude and elevation angle.

Solar-zenith angle is the angle between the observer's zenith and the sun. The solar-zenith angle dependence of vertical ionospheric electron profile has been modeled by Chapman (Ref. 3) and by Yip, von Roos and Escobal (Ref. 4). Parameters in the SEASAT altimeter semi-empirical model of S. C. Wu (private communication, 1977) applied in this analysis are determined by least-squares fitting to the measured zenith electron content. Daytime zenith electron content is lower during summer than during winter while nighttime zenith electron content remains higher during summer. Thus the model parameters vary during the year.

The geomagnetic adjustment is complicated by a "geomagnetic anomaly." During early morning hours, the concentration of ionospheric electrons is higher at the magnetic equator and lower at the geomagnetic poles. As the ionosphere is illuminated by the sun, electrons drift to the north and south away from the magnetic equator reducing the electron concentration at the magnetic equator by about 10 percent. This time dependence is parameterized in the model of Wu.

The model used to compute ionospheric delay from Faraday rotation data refers to the combination of mapping factors used. In this analysis two models for treating the Faraday rotation data are compared. Model 1 accounts for changes in hour angle, elevation angle, solar zenith angle and geomagnetic latitude. Model 2 accounts for changes in hour angle and elevation only. In either case, the difference in ionospheric delay between VLBI stations is calculated according to Eq. (9) from the mapped total electron content.

IV. Data

The VLBI data used in this analysis are for the 257.6-km baseline between the Owens Valley Radio Observatory (OVRO) near Big Pine, California and Deep Space Station 13 (DSS 13) at the Goldstone Tracking Station near Barstow, California. These data were collected using the Mark III VLBI data acquisition system during Mobile VLBI experiments 81A, 81B and 81C conducted on February 15, 17 and 18 respectively in 1981. Correlation of these experiments was done at the Haystack Observatory, Westford, Massachusetts. Phase calibration was applied in all three experiments. The mean frequencies for the S and X bands were 2289.901 MHz and 8437.9102 MHz, respectively. Typical system error for synthesized delay in these data is 0.05 nanoseconds.

Values of columnar total electron content at intervals of one hour were extracted from Faraday rotation data covering the time period of the VLBI experiments. These data, used in CALIBRATE to interpolate to the mean time of the VLBI observations, are plotted in Fig. 1 in units of 10^{17} electrons per square meter as a function of local time in hours.

For the purpose of comparing the ionospheric delay calculated by the dual frequency and Faraday rotation methods, the delays were averaged over all observations in an experiment and the average was subtracted from the ionospheric delays. Thus, the scatter of the ionospheric delay about the experimental means are the quantities being compared. From Eq. (13), the ratio of ionospheric delay at X-band to that at S-band is roughly 0.08. Thus, S-band ionospheric delays are dominant and they are the data used in this analysis.

In all plots there is one point for each observation. The average time interval between observation mean times is about twenty minutes. A few points in the data had anomalously large delays. Observations resulting in delays outside of the range -3.0 to $+3.0$ nanoseconds were excluded from the analysis. Figures 2, 3 and 4 present the ionospheric delay scatter obtained from both methods plotted against local time for experiments 81A, 81B and 81C, respectively. In the plots against local time, the experiments proceed from the start time of the experiment to 24 hours on the right and are then continued from 0 hours on the left.

V. Discussion

In Fig. 5, the Faraday rotation data for both mapping models are plotted against the dual frequency data. In these scatter plots the data of the three experiments are combined and plotted in the form of deviations from the mean normalized with the standard deviations. Also shown in each figure are the lines of regression, the slopes of which are the correlation coefficients. Since the data are presented in the form of deviations from the mean, the line of regression passes through the origin. The correlation coefficients are 0.23 and 0.47 for models 1 and 2, respectively. Thus, while there is a positive correlation between the dual frequency and Faraday rotation methods of obtaining ionospheric delay, it is a weak correlation. Furthermore, the correlation is made worse rather than better by the inclusion of mapping factors which incorporate solar zenith angle and geomagnetic latitude effects.

In Fig. 6, the data from the three experiments are superimposed in the same plot against local time. In the three experiments, the Faraday rotation data are seen to track very closely, the dual frequency data less well. The same 24-hour schedule was shifted and used for all three experiments. These experiments were consecutive and were performed within a four-day period so that the same sources were being observed at the same local times during each experiment to within a few minutes. The variation of the Faraday rotation data with time of day was similar for the three experiments, so that the kind of tracking seen in the Faraday rotation data of Fig. 6 is expected.

Ionospheric variations may be classified in three categories: large-scale spacial (>500 km), large-scale temporal (>1 h) and small-scale (<500 km or <1 h) variations (Ref. 5). Large-scale effects are modeled in the mapping of the Faraday rotation data. Thus, to a first approximation the large-scale effects may be removed from the dual frequency data by subtracting from it the Faraday rotation data. This difference, plotted in Fig. 7, should retain only the effect that small-scale ionospheric variations and irregularities produce in the dual fre-

quency data by virtue of the fact that in this method ionospheric delay is measured independently at each station. We see, however, that the data of model 1, in contrast to that of model 2, contain a diurnal signature. This suggests that the model 1 mapping function introduces a large-scale effect into the data.

Based on the Faraday rotation data, day hours were taken to be 0700 to 2100 local time. Standard deviations σ_D and σ_N for the day and night periods, respectively, of the data in Fig. 7 are given in Table 1. Let us assume that there is a component of the fluctuation in the measured ionospheric delay which is proportional to the total electron content. This implies

$$\sigma_{ID} = R\sigma_{IN} \quad (13)$$

where σ_{ID} and σ_{IN} are respectively the day and night standard deviations of the ionospheric delay, and R is the ratio of the average total electron content of day to that of night. In Table 1, the ratio σ_D/σ_N is 1.48 and 1.35 for the data based on mapping models 1 and 2, respectively. From the Faraday rotation data, the average total electron content of the ionosphere, in units of 10^{17} electrons per square meter, was found to be 1.88 and 5.62 for night and day, respectively, so that $R = 2.99$. The fact that R is different from σ_D/σ_N may be interpreted to indicate that there is an additional component σ_o of the ionospheric delay scatter which is independent of changes due to sunlight. Thus,

$$\begin{aligned} \sigma_D^2 &= \sigma_{ID}^2 + \sigma_o^2 \\ \sigma_N^2 &= \sigma_{IN}^2 + \sigma_o^2 \end{aligned} \quad (14)$$

From Eqs. (13) and (14) one finds the values of σ_{ID} , σ_{IN} and σ_o listed in Table 1.

Two points can be made about the statistics of Table 1. First, that the scatter in the data associated with model 1 is significantly larger than that associated with model 2, reflecting the weaker correlation of the model 1 Faraday rotation data to the dual frequency data. Second, that σ_o , the component of scatter in ionospheric delay which is constant, is comparable in size to σ_{ID} , the component associated with sunlight. System noise error for these data, based on the number of bits correlated, is approximately 0.05 nanoseconds and is therefore too small to account for σ_o . Possibly the component which is independent of total electron content is introduced during data acquisition and processing. On the

other hand, it may be that the assumption of Eq. (13) is incorrect.

Rays from the ends of a baseline in the direction of an extragalactic source penetrate the ionosphere along parallel lines separated by the baseline distance. For the relatively short baseline of these data, fluctuations in ionospheric delay can be estimated according to

$$\sigma_{\tau} = A(L/10)^{0.75} \tau \quad (15)$$

where L is the distance in kilometers between points in the ionosphere and A is a dimensionless constant (Ref. 5). Since ionospheric delay is proportional to total electron content, the estimate of the standard deviation given by Eq. (15) is proportional to total electron content and is therefore comparable to the quantities of σ_{ID} and σ_{IN} . Using the average values of total electron content for day and night, the average ionospheric delay is computed from Eq. (8) to be 14.4 and 4.8 nanoseconds, respectively, for day and night observations. From the values of σ_{ID} and σ_{IN} given in Table 1, we find the constant A to be 0.0045 and 0.0027 for models 1 and 2 respectively. Callahan finds values of A in the range 0.0040 to 0.0031. This agreement supports the contention that the delay scatter seen in Fig. 7 is due to small-scale variations in the ionosphere.

VI. Conclusions

The weakness of the correlation between ionospheric delay computed by the dual frequency and Faraday rotation methods supports the conclusion of J. M. Davidson (private communication, 1981), obtained for the ARIES Project, that the ionosphere has not been adequately modeled in the mapping function applied to Faraday rotation data to calibrate VLBI data. Ionospheric delay computed from Faraday rotation data depends on the measurement of total electron content along a single ray path through the ionosphere. Ionospheric delay determined by the dual frequency method depends on the difference between the total electron content along two rays through the ionosphere separated by the baseline distance. The dual frequency method therefore incorporates fluctuations due to differences in the ionosphere along the two rays. Since these fluctuations cannot be modeled, the Faraday rotation method cannot be used to compute the ionospheric delay calibration at the level required for processing VLBI data.

In this analysis, one component of fluctuation in ionospheric delay obtained by the dual frequency method is associated with variations in the total electron content of the ionosphere induced by sunlight. The other component is found to be independent of day-night variations in the ionosphere. System noise error is too small to account for the solar independent component.

Acknowledgments

The Faraday rotation data was supplied by H. N. Royden. The author wishes to thank G. E. Lanyi, G. H. Purcell, J. L. Faselow and H. N. Royden for helpful discussions.

References

1. Meaks, M. L., *Methods of Experimental Physics*, Vol. 12, part B. New York: Academic Press, 1976.
2. Thomas, J. B., "An Analysis of Radio Interferometry with the Block O System," *JPL Publication 81-49*, Jet Propulsion Laboratory, Pasadena, California, December, 1981.
3. Chapman, S., "The Absorption and Dissociative or Ionizing Effect of Monochromatic Radiation in an Atmosphere on a Rotating Earth," *Proc. Phys. Soc. (London)*, Vol. 43, p. 25, 1931.
4. von Roos, O. H., Yip, K. B. W. and Escobal, P. R. "A Global Model of the Earth's Ionosphere for Use in Space Applications," *Astronautica Acta*, Vol. 18 (Supplement), p. 215, Pergamon Press, 1974.
5. Callahan, P. S., "Ionospheric Variations Affecting Altimeter Measurements: A Brief Synopsis," *Marine Geodesy*, Vol. 8, Numbers 1-4, 1984.

Table 1. Statistics for dual frequency delay scatter minus Faraday rotation delay scatter. The day and night data includes 107 and 68 data points, respectively.

	Model 1	Model 2
σ_D	0.936 nsec	0.618 nsec
σ_N	0.631 nsec	0.457 nsec
σ_D/σ_N	1.48	1.35
σ_{ID}	0.733 nsec	0.441 nsec
σ_{IN}	0.244 nsec	0.147 nsec
σ_o	0.592 nsec	0.433 nsec
A	0.0045	0.0027

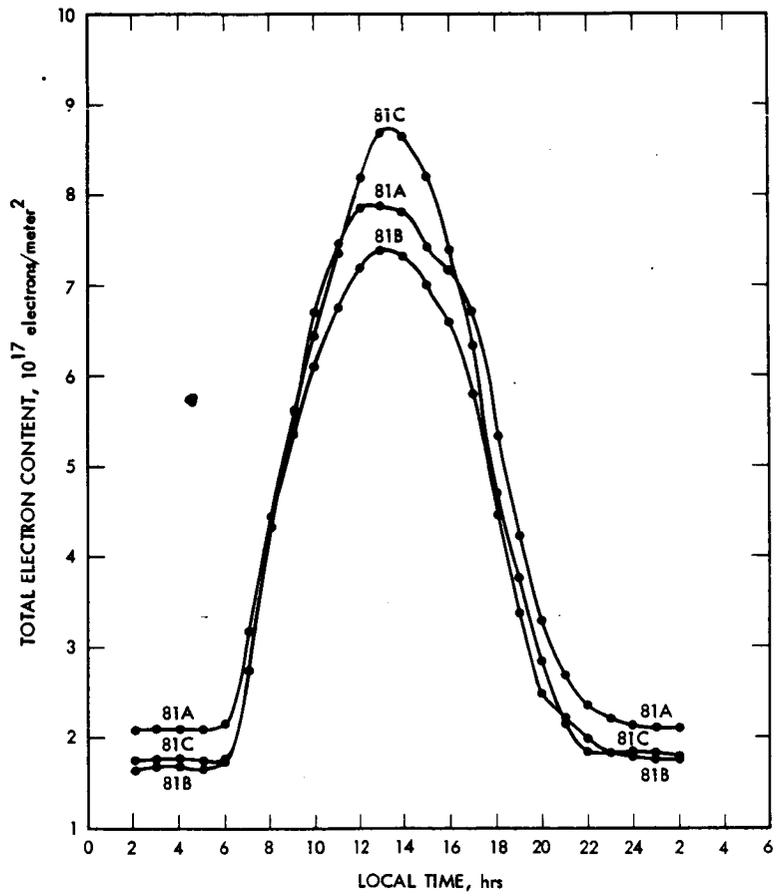


Fig. 1. Columnar total electron content of the ionosphere determined by the Faraday rotation method in units of 10^{17} electrons per square meter versus local time, experiments 81A, 81B and 81C.

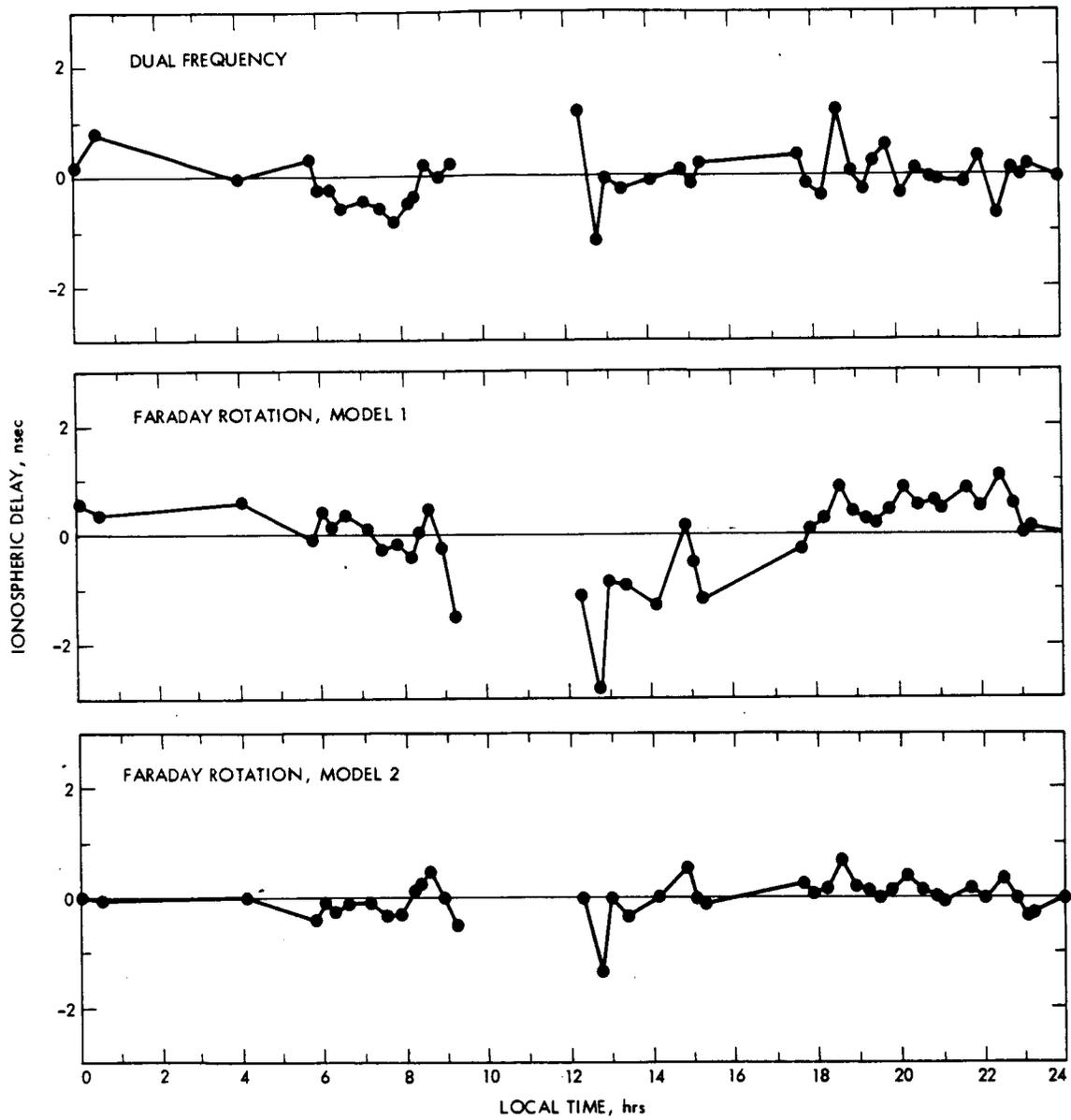


Fig. 2. S-band ionospheric delay scatter versus local time, experiment 81A

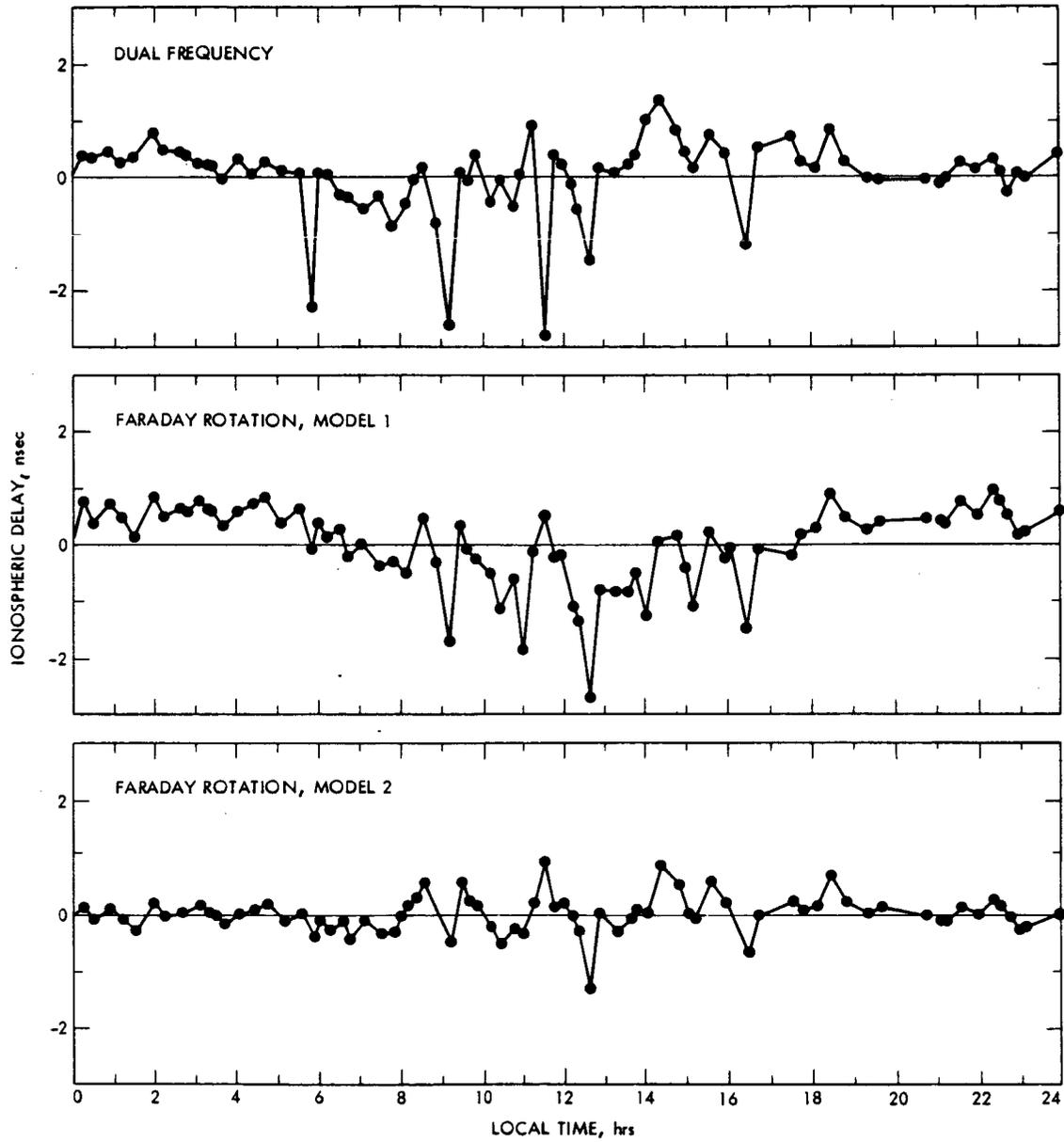


Fig. 3. S-band ionospheric delay scatter versus local time, experiment 81B

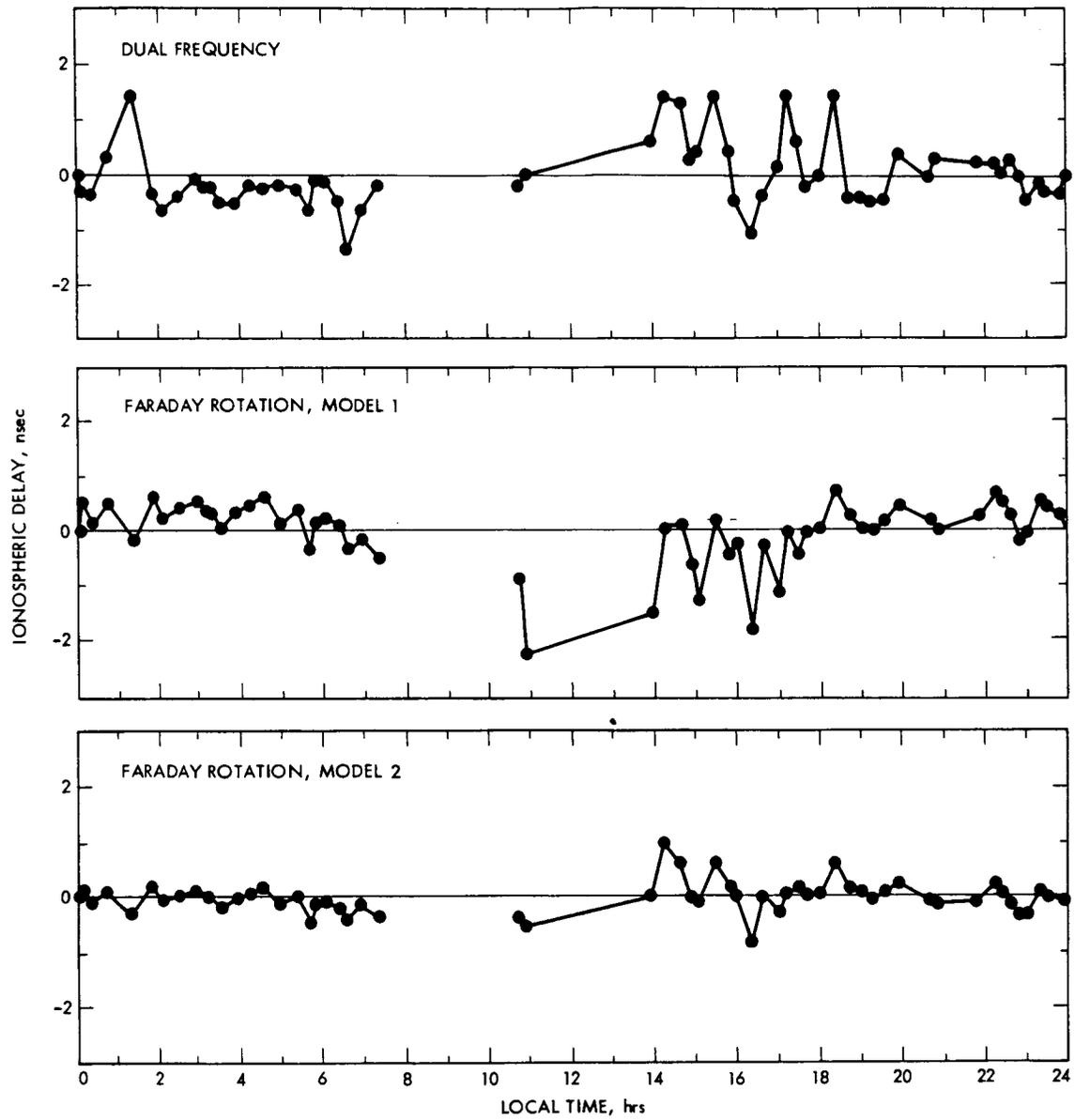


Fig. 4. S-band ionospheric delay scatter versus local time, experiment 81C

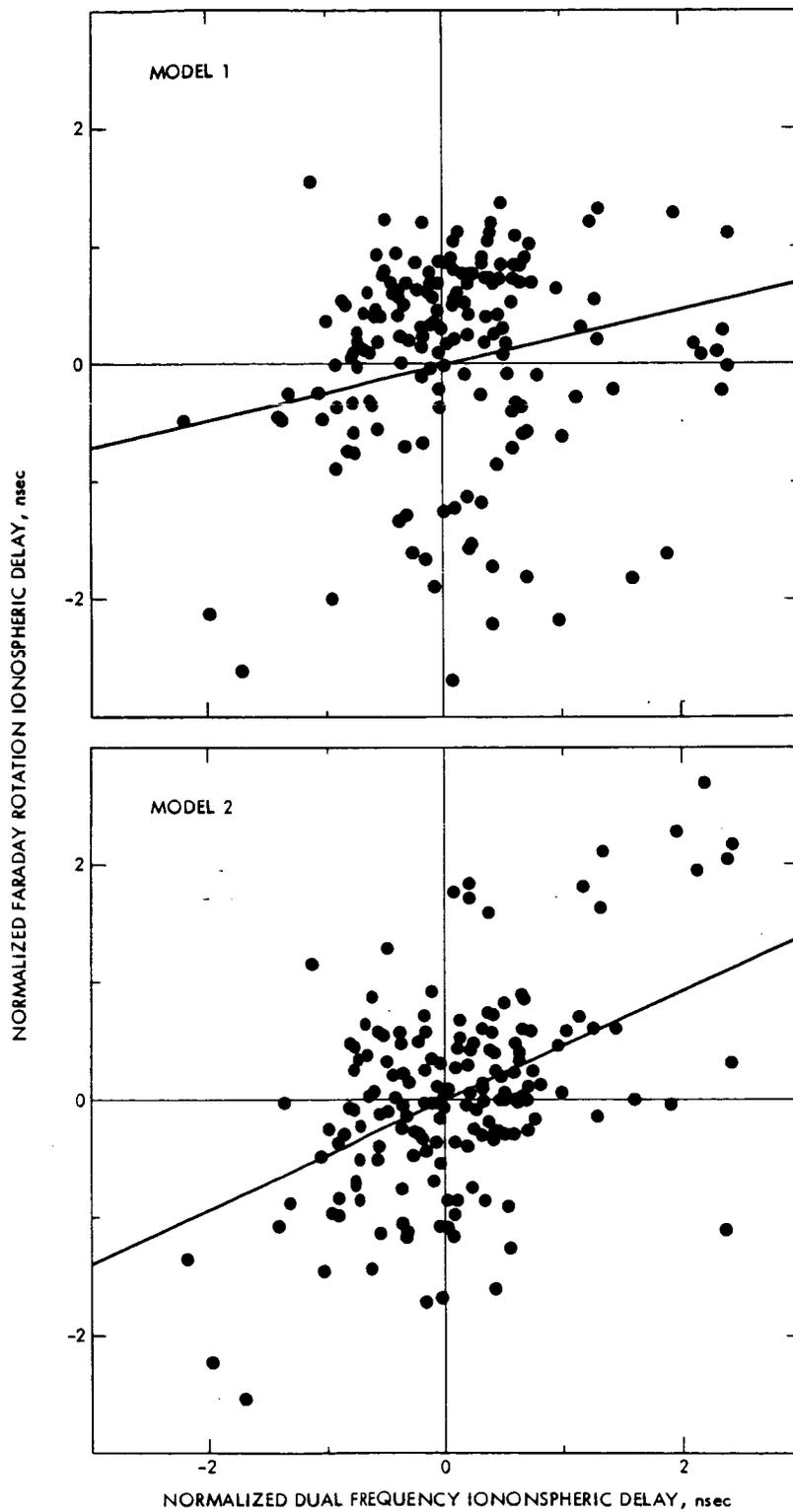


Fig. 5. Dual frequency delay scatter versus Faraday rotation delay scatter at S-band, experiments 81A, 81B and 81C

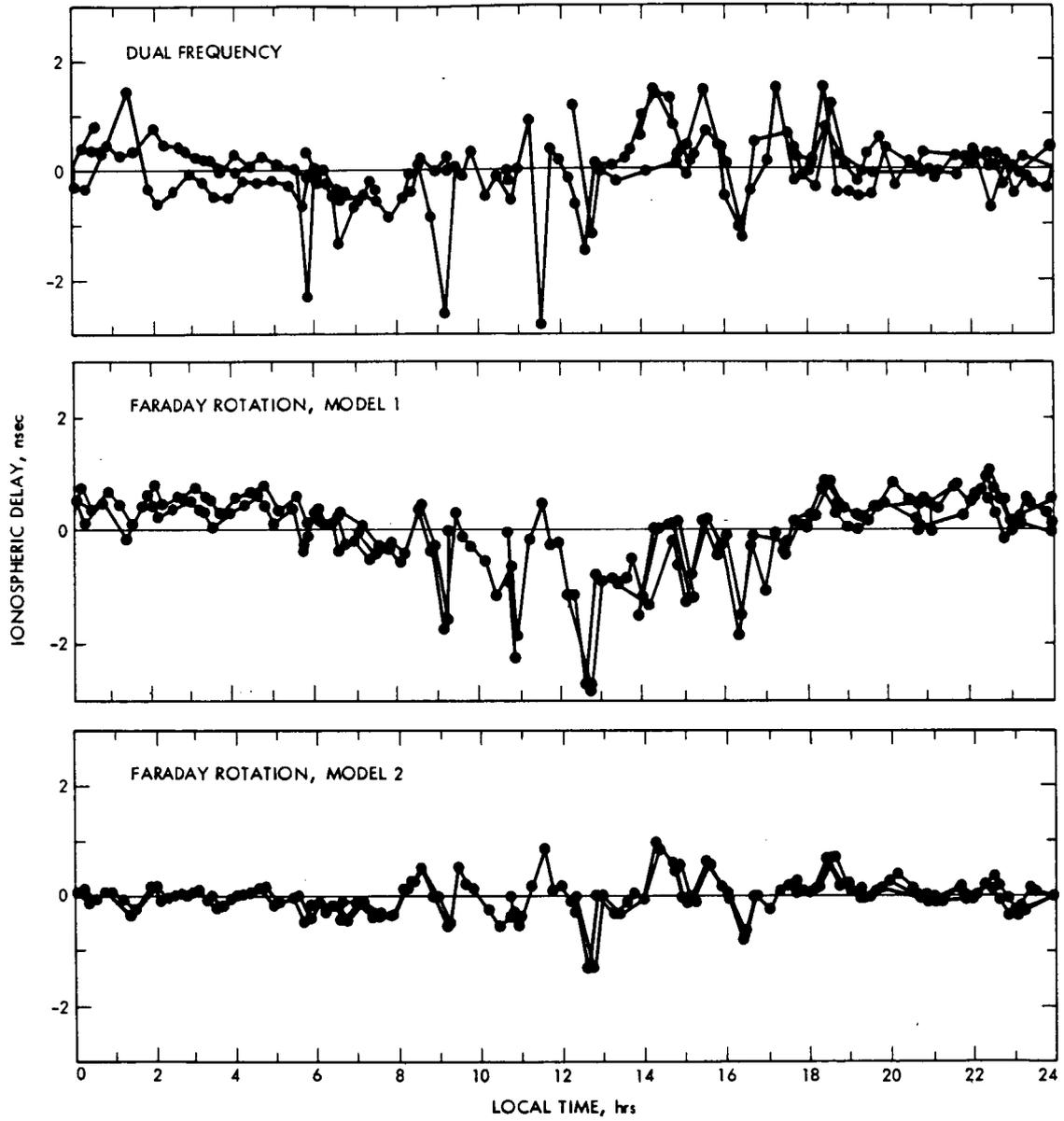


Fig. 6. Dual frequency and Faraday rotation ionospheric delay scatter versus local time at S-band, experiments 81A, 81B and 81C

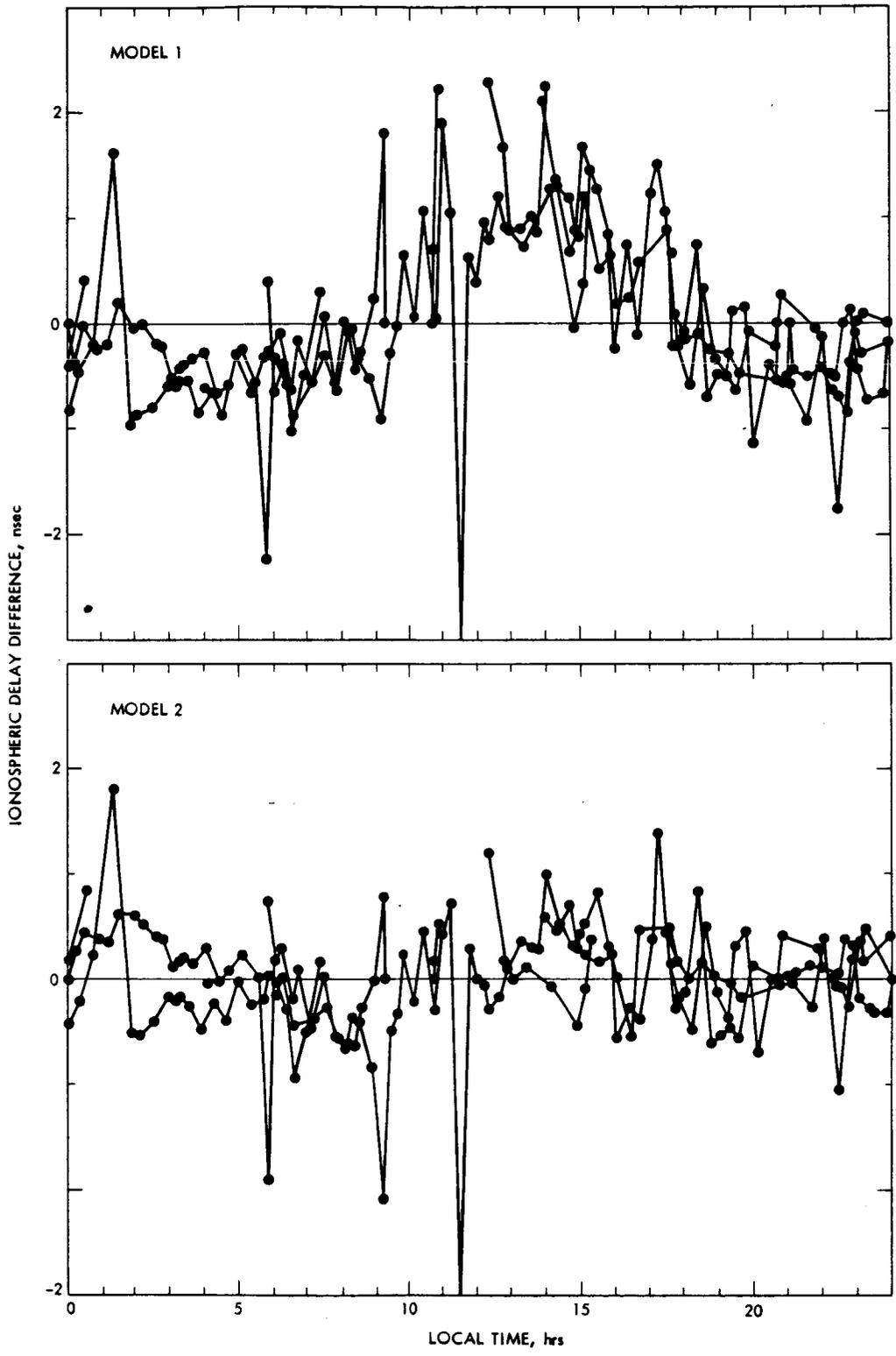


Fig. 7. Dual frequency delay scatter minus Faraday rotation delay scatter versus local time at S-band, experiments 81A, 81B and 81C

A New 70-Meter Antenna Quadripod With Reduced RF Blockage

J. J. Cucchissi

Ground Antennas and Facilities Engineering Section

The new subreflector mount (quadripod) for the 64-meter to 70-meter antenna extension project was the result of many trial designs aimed at reducing RF spherical and plane wave blockage and minimizing structural weight while satisfying strength and natural frequency requirements. An optimum design emerged which has a gain improvement of 0.32 dB over the present 64-meter design. This article describes some of the trial designs made and the final optimum configuration selected.

I. Introduction

The 64-Meter Antenna Rehabilitation and Performance Upgrade Project at the Jet Propulsion Laboratory (JPL) initially aimed at increasing the gain by about 1.9 dB at X-band (8.45 GHz) for the upcoming Voyager-Neptune and Galileo-Jupiter encounters. The effort entails the following:

- (1) Increasing the antenna aperture from 64 to 70 meters (+0.8 dB).
- (2) Resurfacing the entire primary reflector with high quality panels and aligning those panels with high precision optical theodolites and holographic methods (+0.5 dB).
- (3) Shaping the subreflector to an asymmetric surface and the main reflector to an axisymmetric surface in order to obtain a uniform RF radiation pattern (+0.3 dB).
- (4) Automating the axial (z-axis) and y-axis focusing controllers of the subreflector positioner and adding structural stiffening braces to the main reflector central truss to reduce gain loss induced by gravity loads when

the antenna is tilted at extreme elevation angles (+0.3 dB).

- (5) Employing any modifications that may add to the total 1.9 dB gain increase and add a level of confidence to the upgrade project.

Since the subreflector mount was to be redesigned anyway, a vigorous effort, funded jointly by the Advanced Systems Technology and the TDA Engineering offices at JPL, was undertaken to investigate different quadripod configurations that showed a promise in reducing the RF blockage. When compared to the other structural and mechanical modifications, the quadripod yielded the greatest potential gain improvement for its cost. A thorough review of candidate designs was made, and several were selected to be studied.

II. Design Requirements

For each new configuration, the following requirements were in effect:

- (1) The total RF blockage (planar and spherical wave) due to the quadripod legs should be no greater than the percentage aperture area blockage for the current 64-meter antenna. If possible, the blockage should be minimized.
- (2) The natural frequency of the lowest mode of the quadripod should be equal to or greater than that for the present 64-meter antenna quadripod to be compatible with the existing servo system.
- (3) The lateral (y -gravity) displacement of the subreflector should not be excessive due to the limited capability of the subreflector positioner to correct for it (9 in. range).
- (4) The quadripod should conform to the structural strength criteria set forth in the ASCE Tower Code (Ref. 1) under varying gravity loading, corresponding to different antenna elevation angles. Furthermore, the quadripod should accommodate a larger, heavier subreflector and subreflector positioner and be able to withstand safely the lifting loads of the feedcone and subreflector hoists.
- (5) The quadripod should be an "optimum" design rather than just a "working" design, i.e., it should satisfy the above performance constraints with minimum truss weight. Material cost savings are related to weight reduction.

III. Definition of Blockage

The beam of transmitted or received energy includes all rays parallel to the axis of the paraboloid that fall within the aperture of the paraboloid. The beam intensity is assumed to be uniform over its circular cross-section. The blockage area is defined as that portion of the beam cross-section representing the optically obstructed rays (Ref. 2). It can be considered as the sum of two kinds of blockage: (1) the blockage that occurs where the wavefront is spherical, and (2) the blockage that occurs where the wavefront is plane. Spherical wave blockage is the shadow of the quadripod legs projected on the paraboloid when they are illuminated from the focus. In Fig. 1, this is the shaded area outside the dashed circle. Plane wave blockage is the projection of the subreflector and quadripod onto a plane and corresponds to the shaded area inside the dashed circle.

Without an explicit function to relate optical blockage to the RF antenna efficiency, the following empirical equation valid for small B was used to relate the percent of blocked area to aperture area, B , and aperture efficiency, η :

$$\eta = [1 - 1.2(B/100)]^2$$

For the purpose of this study, the quadripod legs were assumed 100% opaque. Since the inner face of the legs would be constructed from steel plates to simplify fabrication, this assumption is not only conservative but accurate. Also, gain loss due to the subreflector plane wave blockage is not included in the gain loss values given in Tables 1 and 2, because its effect is usually accounted for in the microwave efficiency estimate. The "net blockage," as defined here, includes optical plane wave and spherical wave blockage by the quadripod legs only.

IV. Methodology

Three different options to reduce RF blockage by the quadripod are indicated in Fig. 2. In order of increasing reduction potential, they are: (1) increase $R1$ by attaching the legs to the main reflector surface at a point farther from the paraboloidal axis; (2) increase the pitch angle β to make the legs closer to the vertical; and (3) change the cross-section of the legs by reducing the widths of the inner and outer faces, W_i and W_o . The first option would have required extensive modification of the existing main reflector rectangular girder structure to attach the quadripod. Excessive fabrication and erection costs caused this option to be eliminated. The second option reduces spherical wave blockage by moving the inner face of the legs farther from the primary focus of the Cassegrainian system. It also increases the size of the apex, W_a , but this had to be done anyway to accommodate the larger subreflector. The magnitude of the pitch angle was restricted, however, to that necessary for reasonable clearance of the subreflector. The third option became the key to the study and proved to be quite effective.

Using the JPL-IDEAS structural optimization computer program to perform analysis and design, a series of pin-jointed truss finite element models of candidate designs was generated. The performance of the various quadripod geometries was determined for each of the required loading conditions. In each case the member connectivity (or topology) was maintained while the dimensions or proportions of the quadripod leg cross-sectional width profile (Fig. 2, Sec. A-A) were changed. The widths of the parallel faces were selected so that they were "balanced," that is, the outer face (W_o) lay within the spherical wave shadow generated by rays impinging on the inner face (W_i). Minor changes in apex dimensions were made as necessary to match the width changes. The selection of a trapezoidal cross-section minimizes blockage while maximizing quadripod leg torsional stiffness, which is a function of the enclosed cross-sectional area. Also, due to the symmetry of the structure, it was possible to use only half models; this reduced computer computation costs.

V. Trial Designs

Table 1 compares the 64-meter antenna quadripod finite element model with the four best 70-meter trial configurations. All were pin-jointed truss models. For each trial design case, the objective was to meet or exceed the first mode (torsional) frequency of the 64-meter quadripod model while minimizing the structure weight. The gravity displacement of the subreflector was monitored but was not usually imposed as a design constraint.

As the cross-section of the leg was reduced, so was the net blockage. As shown in the table, this reduction is accompanied by a decrease in natural frequency and an increase in gravity displacement of the subreflector. The progressive drop in weight in Models 1, 2, and 3 resulted from a reduction in length of certain leg truss members as well as the relative ease with which the structure met the design constraint.

Model 4 represents a different situation. The increase in structure weight indicated a difficulty in meeting the constraint with a very slender quadripod leg profile. Nevertheless, the weight increase was within acceptable limits, and Model 4 did meet the design constraint; it was selected as the final configuration because of its low blockage.

The "acceptable weight" limitation was not so much a requirement as a guideline. Since these were stand-alone quadripod models, it was not possible to determine directly their effect on the reflector back up truss structure without costly analyses of the larger tipping structure model. Also, the quadripod would be counterweighted, and available space for counterweight was limited. Nevertheless, subsequent analysis of the composite back up structure with quadripod computer models showed no significant penalties resulting from this design.

VI. Final Design

Trusses are usually analyzed as pin-jointed structures. In reality they are not pin-jointed because of bending and torsional stiffness at the rigid corners, but axial forces in the truss members predominate so bending and torsional moments are often neglected. An analysis of a rigid-joint 64-meter quadripod model, however, showed a substantial increase in torsional natural frequency (from 0.74 Hz to 1.22 Hz), whereas a rigid-joint 70-meter model did not (from 0.78 Hz to 0.82 Hz). To achieve the higher 64-meter frequency, the final configuration had to be redesigned.

Primarily a computer program redesign sequence consists of resizing truss members without altering the overall propor-

tions or dimensions of the model. Unfortunately, the topology of Model 4 made it unreasonable to attempt to meet this higher natural frequency constraint by only resizing members; the optimization algorithm in IDEAS, in achieving its best solution, reached a point where the effect of additional structural stiffness was offset by the associated increase in structural mass. At this point, the natural frequency had been maximized and could only be increased by reducing nonstructural weight, which was invariant.

Because of an unwillingness to abandon this configuration (because of its blockage reduction potential), a new modification was proposed: outrigger braces, as sketched in Fig. 3, were added to connect the lower portion of the quadripod legs with additional points on the reflector back up truss. These braces resist the rotation of the legs about their longitudinal axis (which was evident from an examination of the first mode shape) and thereby increase the frequency. The results are listed in Table 2. Compare the results to Model 4 in Table 1, and note the reduced quadripod structural weight as an additional benefit. An insignificant increase in net blockage is due to the outriggers.

VII. Supplementary Results

Additional observations, resulting from other analyses performed during the course of the study, are briefly discussed below:

- (1) For a given quadripod geometry, increasing the aperture area while keeping other parameters constant causes an increase in spherical blockage area, and the percentage increase in spherical blockage is much greater than the percentage increase in aperture area.
- (2) Varying the leg depth had no appreciable effect on either blockage or the frequency of the first mode.
- (3) The first mode shape indicated a rotation of the nearly rigid apex about the focal axis; the legs showed weak-axis bending. An attempt to reduce this bending by using *K*-braces to subdivide the bays of the outer face produced improved but limited results due to the extreme slenderness of the quadripod legs.
- (4) Efforts to meet the torsional frequency requirement included supplementing the apex structure by adding members across the axis of symmetry. This was done early in the study, and results indicated that the connectivity of the apex could be simplified without affecting the first mode frequency. These supplementary members were then removed.
- (5) The second mode of the stand-alone quadripod model exhibited sideways bending at a frequency 0.5 Hz

higher than the first mode. Modes higher than two had frequencies in excess of twice the first mode frequency.

- (6) For the composite back up structure with quadripod model, the lowest mode was quadripod torsion at a frequency slightly less than the first mode of the stand-alone quadripod model.

VIII. Conclusions

The following are conclusions from the quadripod blockage reduction analysis:

- (1) Varying the inner and outer face widths had a significant effect on the quadripod blockage and fundamental frequency.
- (2) The torsional frequency performance constraint controlled the design. In all cases, the requirement on the y -gravity displacement of the subreflector was easily satisfied.

- (3) Experience with the 64-meter quadripod model indicates that a pin-jointed model yields a torsional natural frequency much lower than measured on the actual structure. A rigid-jointed model, which accounts for bending and torsional stiffness of the truss members, gives more realistic results. The 70-meter quadripod model performance was insensitive to the joint continuity. Again, this is probably attributable to the leg slenderness.

- (4) The quadripod legs are usually pin-connected to the reflector truss back up structure. This type of connection cannot resist bending or twisting moments and therefore allows the legs to rotate about their longitudinal axis, which is evident from the mode shape. Limiting this rotation greatly increased the first mode frequency; it required the addition of outrigger braces near the base of the quadripod legs to provide torsional rigidity.

Acknowledgment

The author wishes to express his appreciation to Dr. Roy Levy, Chief Design Engineer of the 70-Meter Extension Project, for his patience and guidance.

Many other individuals have contributed valuable information, time and assistance throughout the course of this study: R. C. Clauss, F. W. Stoller, F. Lansing, H. Phillips, M. S. Katow, S. Rocci, and D. Strain. To these, the author is indebted.

References

1. Guide for Design of Steel Transmission Towers, *ASCE Manual on Engineering Practice*, No. 52, ASCE, New York, 1971.
2. Herndon, J., "Efficient Antenna Systems: A Program to Calculate the Optical Blockage by the Quadripod on Large Microwave Antennas," *JPL Space Programs Summary 37-48*, Vol. II, pp. 58-63, Jet Propulsion Laboratory, Pasadena, CA.

Table 1. Trial designs to reduce RF blockage. All models are pin-jointed.

Parameter	64-m Antenna	70-m Antenna			
		Model 1	Model 2	Model 3	Model 4
W_i , in. (m)	18.0 (0.457)	18.0 (0.457)	18.0 (0.457)	15.0 (0.381)	11.0 (0.279)
W_o , in. (m)	36.0 (0.914)	36.0 (0.914)	32.0 (0.813)	28.0 (0.711)	20.0 (0.508)
H , in. (m)	96.0 (2.438)	96.0 (2.438)	96.0 (2.438)	96.0 (2.438)	96.0 (2.438)
Net Blockage, %	6.34	5.67	5.01	4.44	3.32
Loss, -dB	0.68	0.61	0.54	0.48	0.35
Structure Weight, lb (kg)	41,600 (18,869)	54,258 (24,611)	54,174 (24,573)	53,174 (24,119)	59,146 (26,828)
Gravity Displacement of Subreflector, in. (m)	0.89 (0.0226)	1.25 (0.0318)	1.26 (0.0320)	1.30 (0.0330)	1.48 (0.0376)
Lowest Torsional Frequency, Hz	0.74	1.23	1.13	0.99	0.78

Table 2. Comparison of current 64-m quadripod with final 70-m design. Both models are rigid-jointed.

Parameter	64-m Antenna	70-m Antenna
W_i , in. (m)	18.0 (0.457)	11.0 (0.279)
W_o , in. (m)	36.0 (0.914)	20.0 (0.508)
H , in. (m)	96.0 (2.438)	96.0 (2.438)
Net Blockage, %	6.34	3.42
Loss, -dB	0.68	0.36
Structure Weight, lb (kg)	41,600 (18,869)	53,250 (24,154)
S/R and Positioner Weight, lb (kg)	12,400 (5,625)	24,000 (10,886)
Gravity Displacement of Subreflector, in. (m)	0.89 (0.0226)	1.16 (0.0295)
Lowest Torsional Frequency, Hz	1.22	1.42
Lowest Pitch Frequency, Hz	1.42	2.65

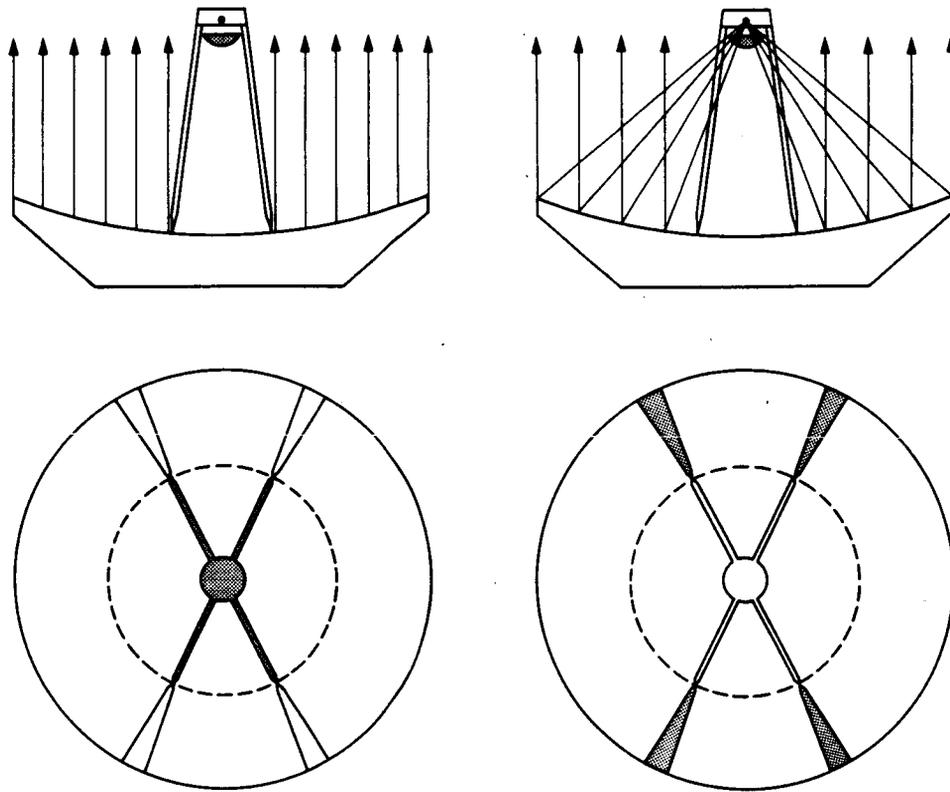


Fig. 1. Plane wave and spherical wave blockage

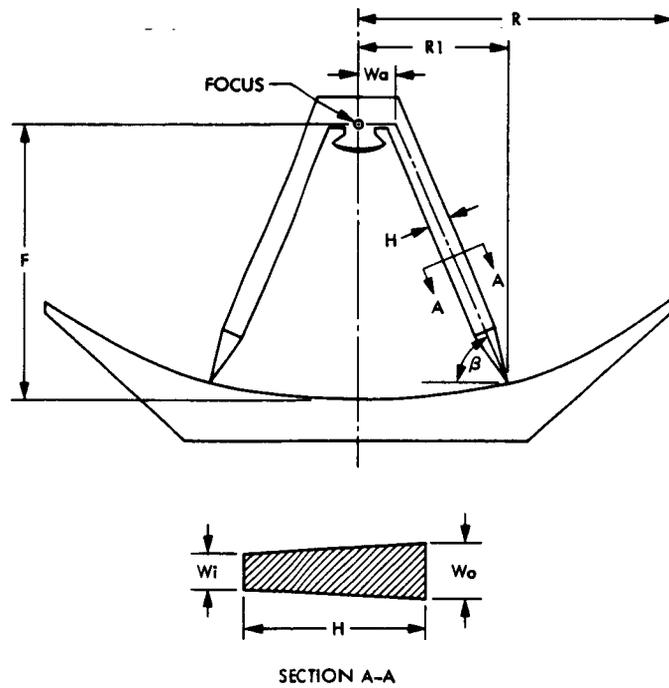


Fig. 2. Parameters controlling RF blockage

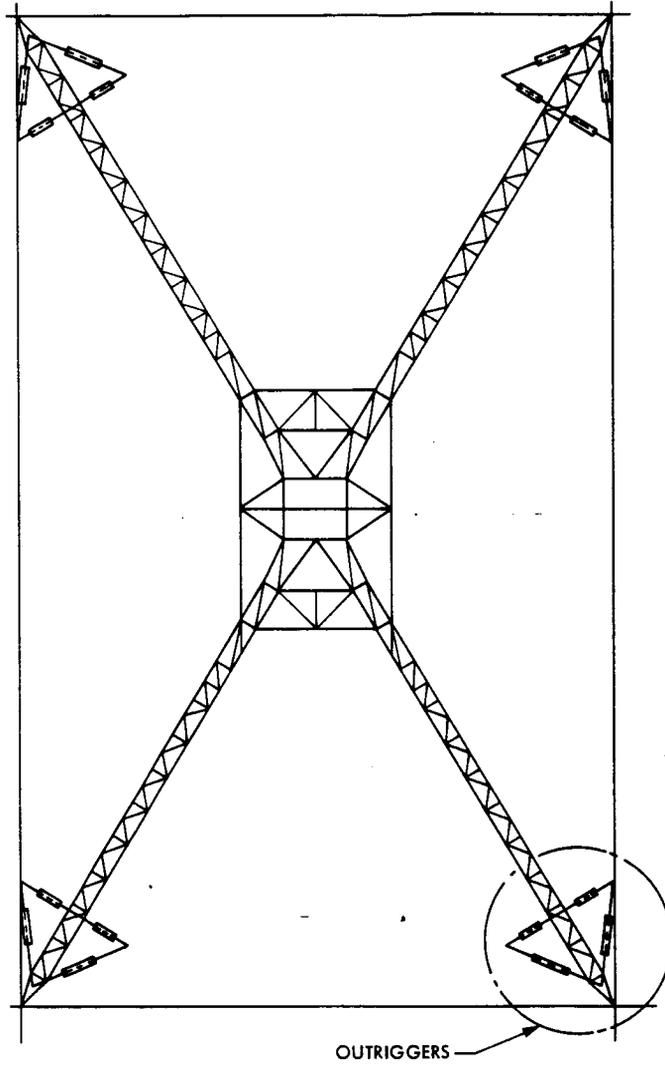


Fig. 3. Plan view of 70-meter quadripod with outriggers

N 86 - 1 0 3 8 6

Seismic Analysis of the Large 70-Meter Antenna, Part I: Earthquake Response Spectra Versus Full Transient Analysis

K. Kiedron and C. T. Chian

Ground Antennas and Facilities Engineering Section

As a check on structure safety aspects, two approaches in seismic analysis for the large 70-m antennas are presented. The first approach, commonly used by civil engineers, utilizes known recommended design response spectra. The second approach, which is the full transient analysis, is versatile and applicable not only to earthquake loading but also to other dynamic forcing functions. The results obtained at the fundamental structural frequency show that the two approaches are in good agreement with each other and both approaches show a safe design. The results also confirm past 64-m antenna seismic studies done by the Caltech Seismology Staff.

I. Introduction

Since the trial structural-mechanical designs and the upgrading and rehabilitation effort of the present 64-m antenna to a 70-m aperture have been completed (Ref. 1), it is essential to check the candidate design for severe environmental loading conditions to satisfy safety requirements. The purpose of this two part study is to investigate the earthquake response of the large antenna structure, to present the modern methodology and to check the structure safety aspects. The emphasis in the first part is placed on the mathematical description of the method of analysis with detailed design safety features to be presented in the second part.

During the early design phase of the 64-m antenna network, the safety against earthquake loading was one of the important tasks that was taken into consideration. The conclusion of that early seismic analysis, conducted by Prof. G. W. Housner of the California Institute of Technology (Ref. 2), was that the structural design should withstand the horizontal acceleration of about $0.25G$ where G is the acceleration of gravity.

Today, with a wider access to modern digital computers and new numerical techniques in structural mechanics, more subtle dynamic analysis of structures is available. This first part of the study will compare the results of two approaches: the response spectra approach versus the full transient analysis. Cross verification of the results should add to a better confidence in understanding the behavior of the complex antenna structure in response to random earthquake excitation.

II. Methodology Description

The analysis of earthquake-excited structures needs to take into account the nonperiodic form of the external forcing term. Such a problem requires the use of special analytical procedures, as presented in Refs. 3, 4, and 5, which are classified into two broad possibilities:

- (1) The frequency response procedure
- (2) The modal analysis procedure

Procedure (1) simply determines the natural undamped frequencies of the structure which the designer uses in comparison against the frequencies of the external forces in an effort to avoid resonance. Procedure (2) is more important and widely used in practice and will be described in this study in detail. The differential equation of the structure under dynamic loading is written in the form:

$$[M] \ddot{y} + [C] \dot{y} + [K] y + \{f\} = 0 \quad (1)$$

where y is generalized displacement, $[M]$, $[C]$, $[K]$ are mass, damping, and stiffness matrices, respectively, and $\{f\}$ is an external force vector where $f(t)$ is an arbitrary function of time. Equation (1) is the basic equation describing the dynamic behavior of the structure.

The first step in solving Eq. (1) is the determination of the free undamped response where no damping or forcing terms exist. The special dynamic form of Eq. (1) is reduced to

$$[M] \ddot{y} + [K] y = 0 \quad (2)$$

The general solution of free responses represented by Eq. (2) is written in the form:

$$y_o = \sum_{i=1}^n y_{oi} \exp(j\omega_i t) \quad (3)$$

where ω_i are the eigenvalues (natural frequencies) of the system, y_{oi} are the eigenvectors, n is the degree of freedom and $j = \sqrt{-1}$.

For forced responses represented by Eq. (1) the solution can be written also in a linear combination of modes as

$$y = \sum_{i=1}^n y_{oi} z_i(t) = [y_o] \{z\} \quad (4)$$

where the matrix $[y_o]$ lists all the modes (assumed to be normalized) and $z_i(t)$ are scalar mode participation factors, $i = 1, \dots, n$, $z_i(t)$ is a function of time and represents the proportion of motion in each mode.

An important advantage of the linear mode superposition in Eq. (4) is that an approximate solution can be obtained by truncating modes and including only part of the total modal contributions. In general, lower modes make the principal contribution to the dynamic response, and good approximation is usually obtained by considering only the first few modes in the analysis.

If Eq. (4) is substituted into Eq. (1) and the result is premultiplied by $[y_o]^T$, then

$$[y_o]^T [M] [y_o] \{\ddot{z}\} + [y_o]^T [C] [y_o] \{\dot{z}\} + [y_o]^T [K] [y_o] \{z\} + [y_o]^T \{f\} = 0 \quad (5)$$

By the orthogonality property,

$$\{y_{oi}\}^T [M] \{y_{oj}\} = \begin{cases} 0 & i \neq j \\ M_i^* & i = j \end{cases} \quad (6)$$

where $i, j = 1, \dots, n$ and M_i^* is a generalized mass.

Also by definition of the eigenvalue problem,

$$[K] \{y_{oi}\} = \omega_i^2 [M] \{y_{oi}\}$$

then

$$\{y_{oi}\}^T [K] \{y_{oj}\} = \begin{cases} 0 & i \neq j \\ \omega_i^2 M_i^* & i = j \end{cases} \quad (7)$$

The matrix $[C]$ is of such a form that

$$\{y_{oi}\} [C] \{y_{oj}\} = \begin{cases} 0 & i \neq j \\ 2 \omega_i \xi_i M_i^* & i = j \end{cases} \quad (8)$$

where ξ_i is the damping ratio. The natural frequency of a structure affects its response to an earthquake. Most of the energy content of an earthquake is in the 1 to 20 Hz frequency range. The duration of "violent" shaking may last 20 or more seconds. If a structure has a fundamental natural frequency in the 1 to 20 Hz range, then it will have time to build up a resonant response. The level of resonance built up depends on the structural damping. For antenna structures considered here, a damping ratio of 7% for the concrete and 4% for the steel is reasonable. Table 1 lists recommended damping ratio values (Ref. 6). For parameterization, we took damping ratios as 2, 5 and 10% as described later in Section III.

Then, the system of Eq. (5) contains only diagonal terms and forms simply a set of ordinary differential equations which, after the modes are normalized, are written as

$$\left. \begin{aligned} \ddot{z}_1 + 2 \xi_1 \omega_1 \dot{z}_1 + \omega_1^2 z_1 &= - \{y_{01}\}^T \{f\} / M_1^* \\ \ddot{z}_2 + 2 \xi_2 \omega_2 \dot{z}_2 + \omega_2^2 z_2 &= - \{y_{02}\}^T \{f\} / M_2^* \\ \vdots & \\ \ddot{z}_n + 2 \xi_n \omega_n \dot{z}_n + \omega_n^2 z_n &= - \{y_{0n}\}^T \{f\} / M_n^* \end{aligned} \right\} (9)$$

Each ordinary differential equation in Eq. (9) can be solved in an elementary manner, and the complete solution is obtained by superposition as in Eq. (4). In the earthquake case, forces $\{f(t)\}$, at each node, vary in the same manner with time. When the antenna is subjected to earthquake motions, the foundation is subjected to a certain "forcing" acceleration $a(t)$, which tends to move with the ground. Since the motion is relatively rapid, it causes severe stresses and deformation throughout the antenna structure. If a mechanical-structural component of the antenna is rigid, it will move with the same acceleration motion of its base, and the dynamic forces acting on it will be very nearly equal to those associated with the base acceleration. By superposing and opposing motions on the whole foundation-to-top structure, we can consider the moving foundation as equivalent to a fixed foundation with forces $(-Ma\{r\})$ acting on the nodes as shown in Fig. 1. The vector $\{r\}$ is an influence vector (Ref. 3) which geometrically connects the acceleration at nodes. The $\{r\}$ consists of ones and zeros only.

A typical modal differential equation from Eq. (9) (for $i = 1, \dots, n$) is written as

$$\ddot{z}'_i + 2 \xi_i \omega_i \dot{z}'_i + \omega_i^2 z'_i = a(t) \quad (10)$$

with

$$z_i = \alpha_i z'_i$$

and

$$\alpha_i = \{y_{oi}\}^T [M] \{r\} / M_i^* \quad (11)$$

The solution of Eq. (10) can be written simply as the Duhamel integral

$$z'_i = \frac{1}{\omega_i} \int_0^t a(\tau) \exp[-\xi_i \omega_i (t - \tau)] \sin \omega_i (t - \tau) d\tau \quad (12)$$

In practice, two approaches to solve Eqs. (10) through (12) are followed:

Approach 1: Full transient analysis. The complete transient response analysis is obtained from integration of Eq. (12), carried out numerically. In principle, superpositions in Eq. (4) will result in the full transient responses required. Often a simple calculation is carried out for each mode to determine maximum responses followed by a suitable "addition" of these responses. More details are given in Appendices A and B.

Approach 2: Response spectra. Direct earthquake response spectra are obtained without the necessity of carrying out complete transient analysis. For various input earthquake motions, the responses of a single degree of freedom (typical of Eq. [10]) have been evaluated to determine the "envelope" response spectra. These are available in the literature and are known as the earthquake response spectra, as in Fig. 2 and 3. The response spectra are useful to design engineers because they embrace the spectra of many observed earthquakes. A structure designer can safely select these design response spectra as inputs that describe the statistically justified excitation of the ground at a given site in the United States. They included three x, y, z motions (two horizontal and one vertical).

III. Computational Results

Twenty values of natural frequencies and the participation factors from Eq. (11) are made available from the JPL IDEAS program for the 70-m antenna. The two approaches are compared for the natural frequency at the first mode, with three different damping ratios 2, 5 and 10%. Table 2 represents the value of permissible displacement S_d and acceleration S_a taken from spectrum curves proposed by Housner in Refs. 7 and 8. The first row in Table 2 concerns an earthquake of the intensity expected at Goldstone, the second row an earthquake of the 1940 El Centro intensity. The third row represents values from the full transient analysis. The computer program TRANST (see also Appendix A) was used to solve Eq. (10) representing the motion of a single degree of freedom of the system. The program solves optionally the eigenvalue problem before proceeding to the transient response solution. The fundamental frequency $f = 1.59$ Hz was used first for computation. A "simplified" 1940 El Centro earthquake was selected as the acceleration input function $a(t)$ in Eq. (10) (Fig. 4). Transient displacement responses are shown in Fig. 5 for damping ratios 2, 5 and 10%, respectively.

Another numerical method was also used to evaluate the Duhamel integral Eq. (12) representing the transient response solution of Eq. (10). This numerical method uses trigonometric identity and converts the original Duhamel integral into a summation of closed form solutions. A description of this method, as well as a flow chart of the computer program are

given in the Appendix B. The transient response for the fundamental frequency $f = 1.59$ Hz, with damping ratios 2, 5 and 10% are given in Fig. 6. The results obtained from both numerical methods are in excellent agreement.

IV. Conclusions

The development of methods and practices suitable for structural design of the DSN large antennas should include the analysis of the structure for earthquake resistance. In this study two approaches were presented, one which utilizes known recommended design response spectra, and the second

which is the full transient analysis, applicable also for a different type of dynamic loading. The comparative results obtained at the fundamental frequency show that the two approaches are in good agreement. The preliminary results agree with past 64-m antenna study done by Caltech Seismology Staff that shows that the center of the mass of the structure should not exceed about 0.25 to 0.35G. In this first part of the study the emphasis was on the mathematical tools to solve the responses and on the cross verification of different approaches. The second part of this study will compute the forces developed in antenna structure components due to seismic excitation and will compare these with seismic design requirements according to building codes.

Acknowledgement

The authors acknowledge the assistance given by Dr. F. L. Lansing, Dr. R. Levy, W. D. Merrick, B. Saldua, D. Strain, J. Cucchissi of JPL and Prof. G. W. Housner of the California Institute of Technology during the various execution steps of this work.

References

1. McClure, D. H. and McLaughlin, F. D., "64-Meter to 70-Meter Antenna Extension," *TDA Progress Report 42-79*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 160-164.
2. TDA Technical Staff, The NASA/JPL 64-Meter-Diameter Antenna at Goldstone California: Project Report, *JPL Technical Memorandum 33-671*, Jet Propulsion Laboratory, Pasadena, California, July 15, 1974.
3. Clough, R. W. and Penzien, L., *Dynamics of Structures*. New York: McGraw-Hill, 1975.
4. Wiegel, R. L., *Earthquake Engineering*. New York: Prentice-Hall, Inc., 1970.
5. Zienkiewicz, O. C., *The Finite Element Method in Engineering Science*. New York: McGraw-Hill, 1977.
6. Newmark, N. M., and Hall, W. J., *Procedures and Criteria for Earthquake Resistance Design in Building Practices for Disaster Mitigation*, Building Science Series 46, U.S. Dept. of Commerce, National Bureau of Standards, pp. 209-236, Feb. 1973.
7. Housner, G. W., "Behavior of Structures During Earthquakes," *Proc. ASCE 85 (EM4)*, 1959, pp. 109-129, Proceedings Paper 2220.
8. U.S. Atomic Energy Commission (1963). *Nuclear Reactors and Earthquakes*, TID-7024, Washington, D.C., Office of Technical Services.

Table 1. Recommended damping ratios

Stress level	Type and condition of structure	Percentage of critical damping
Working stress, no more than about 1/2 yield point	Vital piping	0.5 to 1.0
	Welded steel, prestressed concrete, well reinforced concrete (only slight cracking)	2
	Reinforced concrete with considerable cracking	3 to 5
	Bolted and/or riveted steel, wood structures with nailed or bolted joints	5 to 7
At or just below yield point	Vital piping	2
	Welded steel, prestressed concrete (without complete loss in prestress)	5
	Prestressed concrete with no prestress left	7
	Reinforced concrete	7 to 10
	Bolted and/or riveted steel, wood structures, with bolted joints	10 to 15
	Wood structures with nailed joints	15 to 20

Table 2. Displacement and acceleration for fundamental frequency

$$\omega = 9.99 \text{ rad/s}$$

$$f = 1.59 \text{ Hz}$$

$$T = 0.629 \text{ s}$$

Damping ratio, %	Displacement S_d , in. (cm)	Acceleration, S_d/G
2	1.4 (3.556)	0.362
	2.4 (6.096)	0.620
	3.05 (7.747)	0.826
5	0.9 (2.286)	0.233
	1.56 (3.963)	0.403
	2.76 (7.01)	0.747
10	0.65 (1.651)	0.168
	0.96 (2.438)	0.248
	2.31 (5.867)	0.260

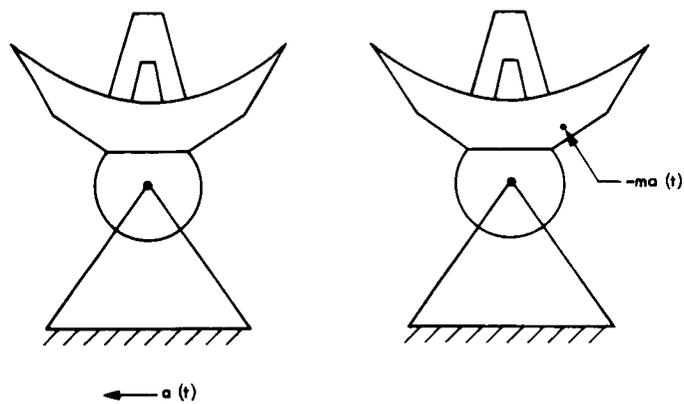


Fig. 1. Foundation motion as equivalent force

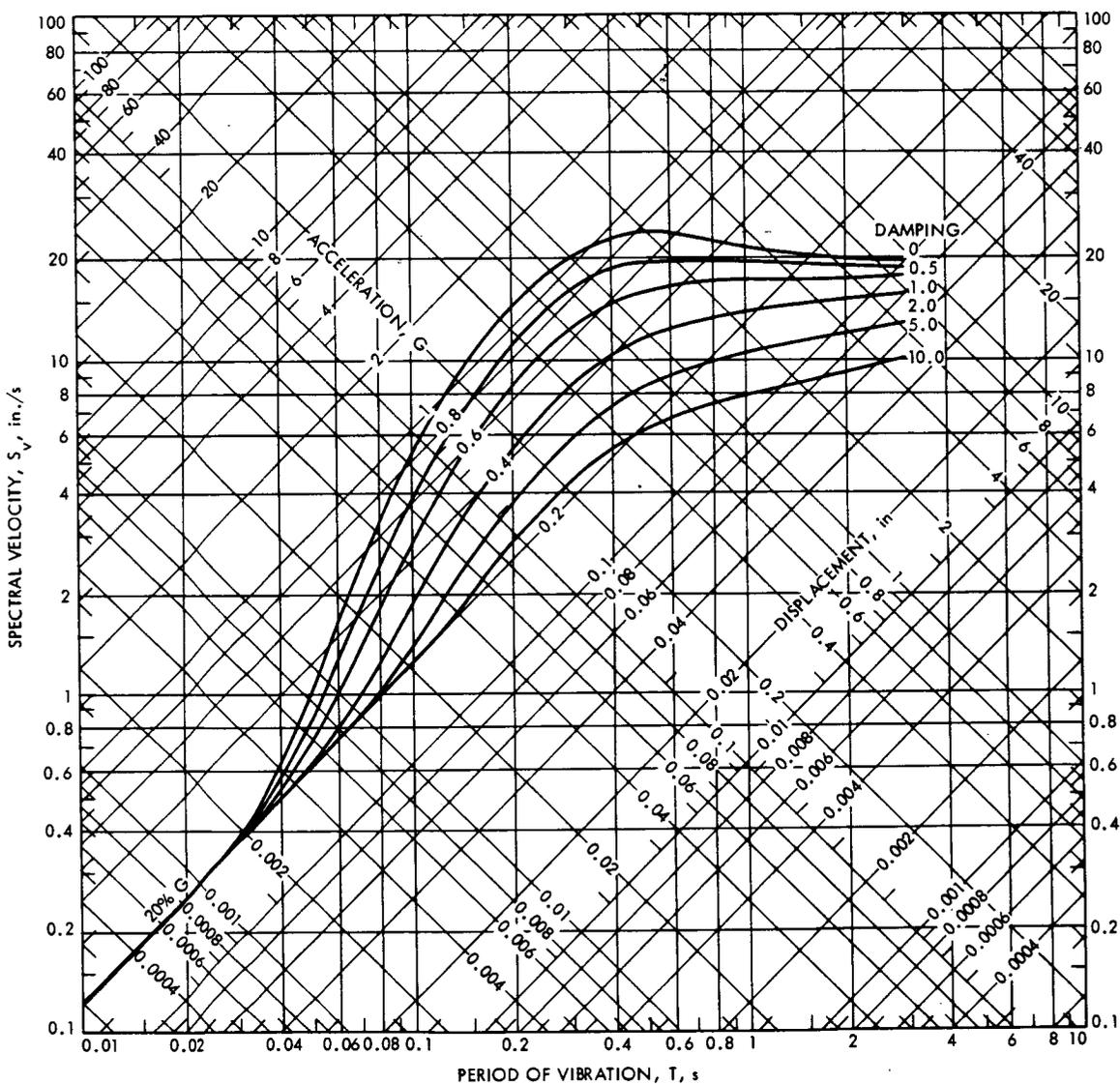


Fig. 2. Combined earthquake response spectra (adapted from Ref. 3)

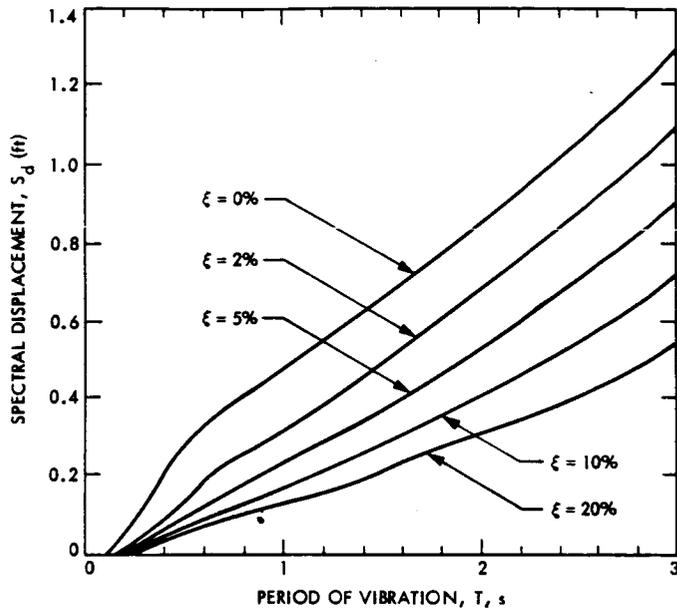


Fig. 3. Average displacement response spectrum of 1940 El Centro intensity (adapted from Ref. 3)

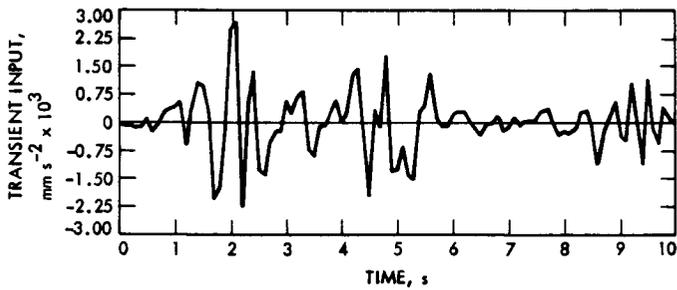


Fig. 4. Earthquake acceleration input function

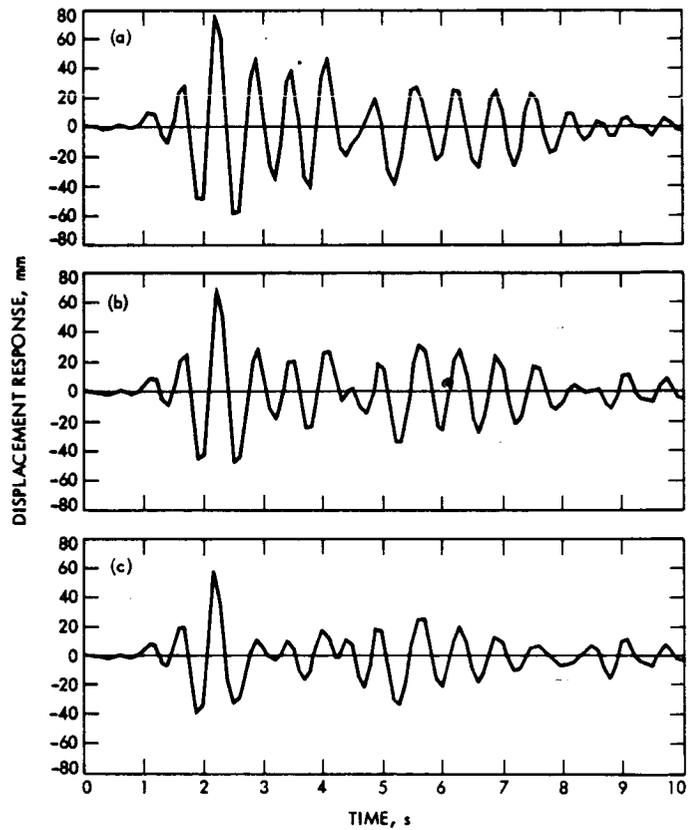
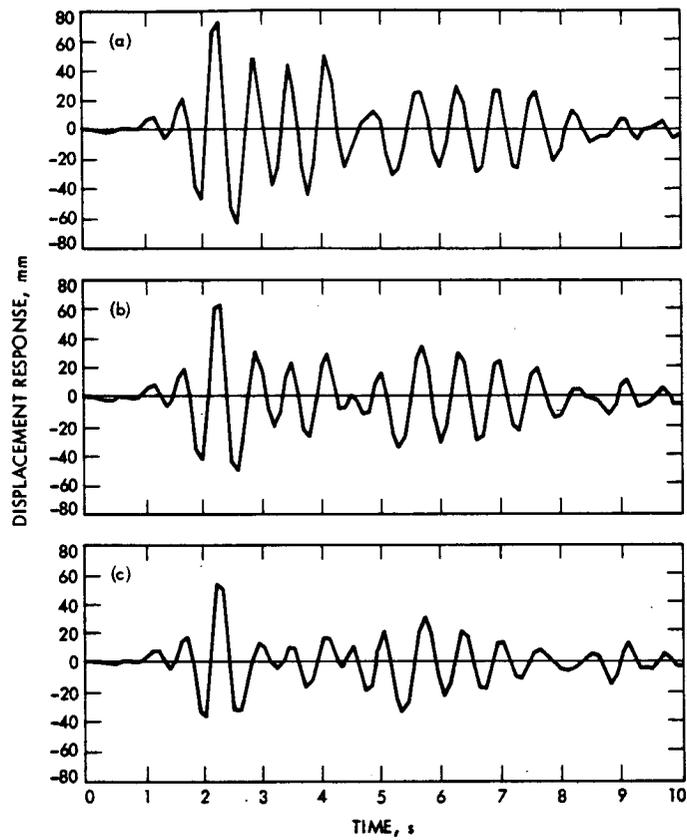


Fig. 5. Transient displacement response damping ratio: (a) 2%, (b) 5%, and (c) 10%



**Fig. 6. Numerical evaluation of Duhamel integral for damping ratio:
(a) 2%, (b) 5%, (c) 10%**

Appendix A

Adams-Moulton Method

The first numerical solution of the antenna transient response follows the Adams-Moulton method in solving a system of linear differential equation of motion, written in the matrix form as:

$$[M] \{\ddot{Z}\} + [C] \{\dot{Z}\} + [K] \{Z\} = \{a(t)\} \quad (\text{A-1})$$

where the matrices $[M]$, $[C]$, and $[K]$ are of the order $N \times N$.

Equation (A-1) can be written as

$$\{\ddot{Z}\} + [M]^{-1} [C] \{\dot{z}\} + [M]^{-1} [K] \{Z\} = [M]^{-1} \{a(t)\} \quad (\text{A-2})$$

assuming $[M]^{-1}$ exists.

Equation (A-2) can be written as

$$\{\ddot{Z}\} = - [M]^{-1} [C] \{\dot{Z}\} - [M]^{-1} [K] \{Z\} + [M]^{-1} \{a(t)\} \quad (\text{A-3})$$

Equation (A-3) can be written in the form of

$$\{\dot{Y}\} = [A] \{Y\} + \{B\} \quad (\text{A-4})$$

if we let

$$\{Y\} = \begin{Bmatrix} Z \\ \dot{Z} \end{Bmatrix} \quad (\text{A-5})$$

$$\{\dot{Y}\} = \begin{Bmatrix} \dot{Z} \\ \ddot{Z} \end{Bmatrix} \quad (\text{A-6})$$

$$[A] = \begin{bmatrix} [O] & [I] \\ -[M]^{-1} [K] & -[M]^{-1} [C] \end{bmatrix} \quad (\text{A-7})$$

and

$$\{B\} = \begin{Bmatrix} [O] \\ [M]^{-1} \{a(t)\} \end{Bmatrix} \quad (\text{A-8})$$

where $[A]$ is the coefficient matrix of the order $2N \times 2N$, $\{B\}$ is a vector of dimension $2N$, and $[O]$ and $[I]$ are the $N \times N$ null and identity matrices, respectively.

Equation (A-4) is a set of $2N$ simultaneous first-order differential equations which are solved by using the Adams-Moulton numerical technique (JPL computer library subroutine SVDQ). The technique uses linear multistep predictor-corrector formulas. Such a technique has the advantage that from successive approximations to each value, an estimate of the truncation error is made.

The fourth-order Runge-Kutta method is used to generate the approximate values of the first four points ($n-3$, $n-2$, $n-1$, n), since the local truncation error is of order h^5 . Values at these previous four points are needed to predict or correct the value at the point ($n+1$). The integration order is selected in such a way as to maximize the step size and to reduce the computation time, consistent with meeting the requested user accuracy.

In the first-order equation

$$\frac{dy}{dx} = f(x, y) \quad (\text{A-9})$$

integrating between x_n and x_{n+1} :

$$\int_{x_n}^{x_{n+1}} \frac{dy}{dx} dx = \int_{x_n}^{x_{n+1}} f(x, y) \quad (\text{A-10})$$

The Adams-Moulton method, like all predictor-corrector methods, starts by predicting y_{n+1} from an initial value of y_n , and then provides successive improvements of y_{n+1} , or else corrects y_{n+1} before calculating the next step.

The Adams-Moulton method uses the following predictor:

$$y_{n+1} = y_n + \frac{h}{24} (55 f_n - 59 f_{n-1} + 37 f_{n-2} - 9 f_{n-3})$$

$$R = O(h^5) \quad (A-11)$$

and the following corrector

$$y_{n+1} = y_n + \frac{h}{24} (9 f_{n+1} + 19 f_n - 5 f_{n-1} + f_{n-2})$$

$$R = O(h^5) \quad (A-12)$$

where h is the step size and R is the truncation error.

By using the Eq. (A-11) as a predictor and Eq. (A-12) as a corrector, the function $y = y(x)$ is obtained. Further details on the method can be found in Ref. A-1.

Reference

- A-1. Carnahan, B., Luther, H., and Wilkes, J., *Applied Numerical Methods*. New York: John Wiley, 1969.

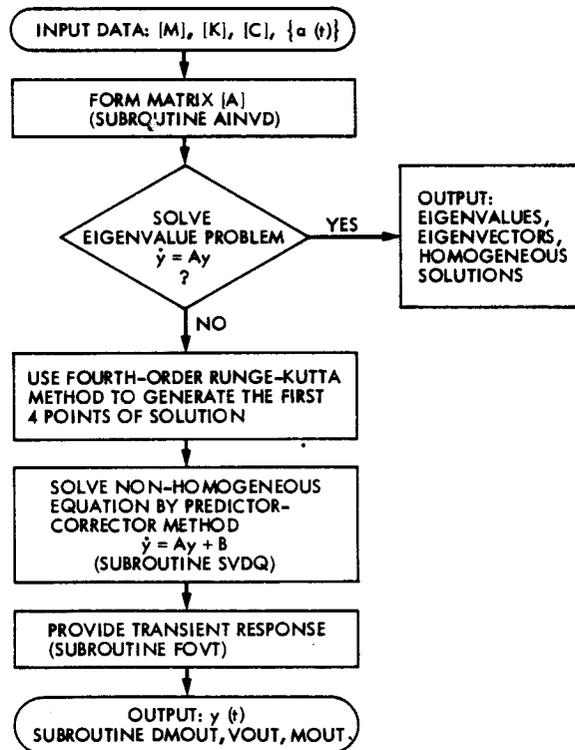


Fig. A-1. Flow chart of TRANST program (using Adams-Moulton method)

Appendix B

Numerical Evaluation of the Duhamel Integral

The Duhamel integral

$$Z'_i(t) = \left(\frac{1}{\omega_i} \right) \int_0^t a(\tau) \exp[-\xi_i \omega_i (t - \tau)] \sin \omega_i (t - \tau) d\tau \quad (\text{B-1})$$

representing the transient response solution of the equation of motion (Eq. [10]) can be evaluated numerically. Since the acceleration input $a(\tau)$ is given in a tabular form, the integrand can be divided into a number of band functions added together in such a manner as to form the original $a(\tau)$ function.

Equation (B-1) is converted into the sum of exact solutions; each resembles Eq. (B-1) but with a unit forcing function $a(t)$. The sum solution can be used when the actual input is approximated by band segments.

The approach is explained as follows: Using the trigonometric relation

$$\sin \omega_i (t - \tau) = \sin \omega_i t \cos \omega_i \tau - \cos \omega_i t \sin \omega_i \tau \quad (\text{B-2})$$

substitute Eq. (B-2) into Eq. (B-1). The Duhamel Integral becomes

$$Z'_i(t) = \frac{\exp(-\xi_i \omega_i t) \sin \omega_i t}{\omega_i} \int_0^t a(\tau) \exp(\xi_i \omega_i \tau) \cos \omega_i \tau d\tau - \frac{\exp(-\xi_i \omega_i t) \cos \omega_i t}{\omega_i} \int_0^t a(\tau) \exp(\xi_i \omega_i \tau) \sin \omega_i \tau d\tau \quad (\text{B-3})$$

Using the relationships

$$\int \exp(\alpha x) \cos \beta x dx = \frac{\exp(\alpha x) (\alpha \cos \beta x + \beta \sin \beta x)}{\alpha^2 + \beta^2} \quad (\text{B-4})$$

and

$$\int \exp(\alpha x) \sin \beta x dx = \frac{\exp(\alpha x) (\alpha \sin \beta x - \beta \cos \beta x)}{\alpha^2 + \beta^2} \quad (\text{B-5})$$

let

$$\alpha = \xi_i \omega_i \quad (\text{B-6})$$

$$\beta = \omega_i$$

and assume that the acceleration $a(\tau)$ is constant (\bar{a}) between $\tau = n \Delta \tau$ and $\tau = (n + 1) \Delta \tau$, where $\Delta \tau$ is the step of summation. Equation (B-3) then becomes

$$Z'_i(t) = \frac{\exp(-\xi_i \omega_i t) \sin \omega_i t}{\omega_i^3 (1 + \xi_i^2)} \times \sum_{\text{all } \Delta \tau} \bar{a} \left[\exp(\xi_i \omega_i \tau) (\xi_i \omega_i \cos \omega_i \tau + \omega_i \sin \omega_i \tau) \right]_{\tau=n \Delta \tau}^{\tau=(n+1) \Delta \tau} - \frac{\exp(-\xi_i \omega_i t) \cos \omega_i t}{\omega_i^3 (1 + \xi_i^2)} \times \sum_{\text{all } \Delta \tau} \bar{a} \left[\exp(\xi_i \omega_i \tau) (\xi_i \omega_i \sin \omega_i \tau - \omega_i \cos \omega_i \tau) \right]_{\tau=n \Delta \tau}^{\tau=(n+1) \Delta \tau} \quad (\text{B-7})$$

for integration step between $\tau = n \Delta \tau$ and $\tau = (n + 1) \Delta \tau$ where n is an integer.

The integral in Eq. (B-1) is then transformed into the summation of the individual contributions from $\tau = 0$ to $\tau = t$. A flow chart of the FORTRAN program used to evaluate Eq. (B-7) is shown in Fig. B-1.

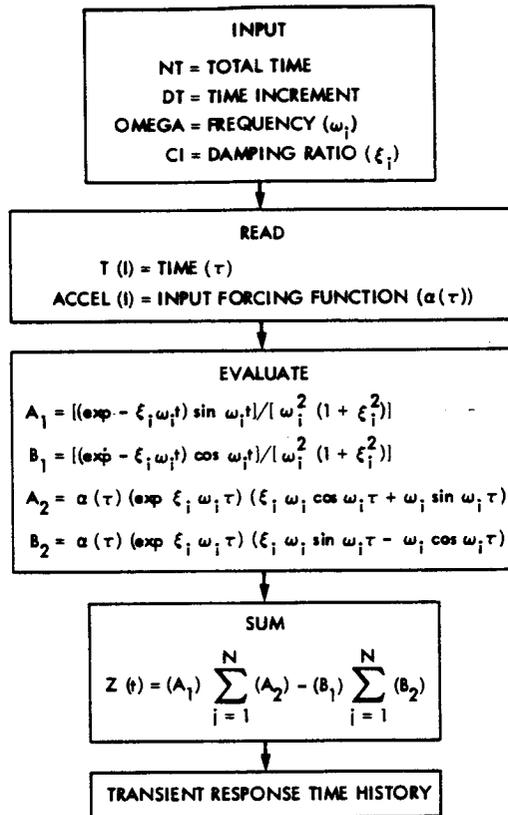


Fig. B-1. Flow chart of numerical evaluation of Duhamel integral

An Experimental $TE_{12} - TE_{11}$ Circular Waveguide Mode Converter

D. Hoppe

Radio Frequency and Microwave Subsystems Section

This article describes theoretical and experimental results for a prototype $TE_{12} - TE_{11}$ circular waveguide mode converter. The system which requires such a device, the high power Ka-Band transmitter is described briefly. A short review of coupled mode theory is given, and the theoretical performance of the final converter design is given. Experimental results for the fabricated converter are presented and compared with theory. A method of identifying the various circular waveguide modes in a multimode device is described. Given the close agreement between the theoretical predictions and experimental results, the computer code may be used with confidence in the design of future multimode tapers and mode converters.

I. Introduction

Previous reports have described the conceptual design of a high power Ka-Band transmitter (Ref. 1), theoretical calculations of mode purity effects on the performance of the system (Ref. 2), and the feed which will be used with the Ka-Band transmitter and transmission line (Ref. 3). This article focuses on another component in the system, the $TE_{12} - TE_{11}$ mode converter.

Conventional klystrons are not capable of producing 400 kW CW at 34.5 GHz, which is the required power level and frequency for the Ka-Band transmitter. The problem arises since, as the klystron output cavity is scaled to higher frequencies, its dimensions are decreased. It soon becomes impossible to extract 400 kW CW from the beam in the reduced interaction volume without exceeding a power density of 1 kW/cm² on the cavity walls. The power density of 1 kW/cm² represents an approximate upper limit which is set by the present state of the art in cooling technology.

In order to increase the interaction volume, an unconventional microwave tube, the gyroklystron, will be used. Two possible configurations for such a device are shown in Fig. 1. In both cases several circular waveguide cavities operating in the dominant TE_{11} circular waveguide mode prebunch the beam. The energy is extracted in the output cavity, which is an open-ended resonator that resonates in the higher order TE_{12} circular waveguide mode. This allows the output cavity dimensions to be increased thus allowing 400 kW to be extracted from the beam without exceeding 1 kW/cm² on the cavity walls. The microwave energy then exits the output cavity in a rotating TE_{12} circular waveguide mode.

Unfortunately the TE_{11} mode, not the TE_{12} mode, is the most suitable mode to use in the rest of the system. The TE_{12} mode has a radiation pattern which is totally unsuitable for illuminating the subreflector. The pattern is multilobed with the main radiation appearing off the waveguide axis. The TE_{11} mode on the other hand has the conventional dominant-mode pattern which is easily modified by the feed to give the

optimum Gaussian pattern for illuminating the shaped sub-reflector. For these reasons it is necessary to develop a device which will convert the TE_{12} mode to the TE_{11} mode. Two different locations for the converter are possible, one inside the vacuum envelope of the tube (Fig. 1[a]), and the other in the 1.75-in. transmission line (Fig. 1[b]). This article describes experimental and theoretical results for a small diameter device suitable for the in-tube configuration depicted in Fig. 1[a].

II. Mode Converter Theory

When the diameter of a circular waveguide exceeds $0.766\lambda_0$ at the frequency of operation, the microwave signal may propagate in more than one circular waveguide mode. For a perfectly straight circular waveguide, these modes are orthogonal and no energy is exchanged between them. When the guide deviates from perfection, either by design or by accident, mode conversion occurs, and the modes become coupled.

The general deformed waveguide may be specified by writing the radius as a function of z , and ϕ as follows (Ref. 4, and J. Doane, "Propagation and mode coupling in corrugated and smooth wall circular waveguide," Plasma Physics Laboratory (internal document), Princeton, New Jersey):

$$r(z, \phi) = r_0 + \sum_{\ell} \alpha_{\ell}(z) \cos \ell\phi + \sum_k \alpha_k(z) \sin k\phi \quad (1)$$

In general, the azimuthal order of the perturbation (ℓ and k) determines which modes will be coupled. For example, a pure $\ell = 1$ perturbation causes coupling between the TE_{ij} , TM_{ij} mode group and the $TE_{(i\pm 1)m}$ and $TM_{(i\pm 1)m}$ mode groups. The $\ell = 1$ perturbation corresponds to curvature. Therefore an incident TE_{11} mode is coupled to the TE_{0m} , TM_{0m} , TE_{2m} and TM_{2m} modes through the curvature. Similarly a radial perturbation with no azimuthal variation, $\ell = 0$, and only a longitudinal variation, couples modes with the same azimuthal index. This is the type of coupling that occurs in circular waveguide tapers and horns.

The first step in a mode converter design is to determine what order of azimuthal variation is required. For example, a $TE_{ij} - TE_{mn}$ converter requires a perturbation of order ℓ where $\ell = |m - i|$. For modes of the same first index a circularly symmetric ($\ell = 0$) perturbation is required, for modes differing by 1 in first index a curvature ($\ell = 1$) perturbation is needed, modes differing by 2 require an elliptical deformation, and so on.

When the $\ell = |m - i|$ perturbation is chosen the incident TE_{ij} mode is coupled to all the $TE_{(i\pm\ell)p}$ and $TM_{(i\pm\ell)p}$ modes. In order to enhance the coupling to only the desired $TE_{(i+\ell)n}$ mode the perturbation is repeated at a specific interval in z , which is approximately given by the beat wavelength between the two modes of interest. The beat wavelength between modes 1 and 2, $\lambda_{1,2}$ is given by

$$\lambda_{1,2} = \frac{\lambda_1 \lambda_2}{|\lambda_1 - \lambda_2|} \quad (2)$$

In summary the rough design of the $TE_{ij} - TE_{mn}$ converter consists of a $\ell = |i - m|$ radial perturbation repeated longitudinally at an interval given by Eq. (2) where mode 1 is the TE_{ij} mode, and mode 2 is the TE_{mn} mode.

In order to accurately determine the number of perturbations required, their magnitude, and their exact placement, a detailed analysis of the coupled mode problem must be undertaken. The propagation in an arbitrarily deformed circular waveguide can be described by the following matrix equation:

$$\frac{d\mathbf{A}(z)}{dz} = -j[\beta(z)]\mathbf{A}(z) + [C(z)]\mathbf{A}(z) \quad (3)$$

Here

$\mathbf{A}(z)$ = a vector containing the mode amplitudes

$[\beta(z)]$ = a diagonal matrix containing the propagation coefficients

$[C(z)]$ = a matrix containing the coupling coefficients for the local waveguide perturbation.

Each term of $[C(z)]$, C_{ij} , is determined by the azimuthal order of the local perturbation, its magnitude, and the specific modes i and j (Ref. 4, and J. Doane, "Propagation and mode coupling in corrugated and smooth wall circular waveguide," Plasma Physics Laboratory (internal document), Princeton, New Jersey). The propagation of coefficient β_{ij} is determined for each mode by using the local radius.

A computer program was developed to solve Eq. (3) for the special case of arbitrary radial perturbations with no azimuthal variation $r(z, \phi) = r(z)$, and modes with first index 1. This is sufficient for the specific mode converter design required, $TE_{12} - TE_{11}$, i.e., $\ell = 0$.

Following the method outlined by Moeller (Ref. 5) the radial perturbation was taken to be sinusoidal with respect to z ,

$$r(z) = r_0 + \Delta r \cos\left(\frac{2\pi z}{\lambda_B}\right) \quad (4)$$

The average radius was chosen to be 0.423 inches which is slightly larger than the radius of the tube's output cavity. The number of ripples was chosen to be seven, and in order to obtain maximum conversion efficiency at the design frequency of 34.5 GHz optimum values of 0.039 inches and 1.364 inches were found for Δr and λ_B respectively. These values were optimized by solving Eq. (3) for various combinations of Δr and λ_B . Seven ripples were chosen since the theoretical ripple magnitude required for the seven ripple device was small enough that no modes become trapped in the ripples at the design frequency of 34.5 GHz. This condition is required for accurate modeling of the interaction with the existing computer code. A cross sectional view of the final device is shown in Fig. 2.

The mode converter is a reciprocal device. That is, an input TE_{12} mode will be converted to an output TE_{11} mode with exactly the same efficiency as an input TE_{11} mode is converted into a TE_{12} mode. Although the final device will be used as a $TE_{12} - TE_{11}$ converter it is simplest to test it reciprocally as a $TE_{11} - TE_{12}$ mode converter. Figure 3 shows the mode composition as a function of z when a TE_{11} mode is incident on the device. In this plot, and throughout the remainder of this paper, dBc denotes the amount of power carried by any waveguide mode with respect to the input TE_{11} mode power. The theoretical final output mode composition is summarized in the last column of Table 1. The efficiency of the conversion is found to be 99.77%. The device efficiency when it is operating in the reciprocal mode will also be 99.77% but the spurious output power will be contained in other modes. The conversion efficiency is plotted as a function of frequency in Fig. 4. The computer program was also used to determine the allowable tolerances on r_0 , Δr , and λ_B to be given to the machine shop in order to ensure high conversion efficiency for the fabricated device. The most sensitive of the parameters was found to be the average radius, r_0 , which must be held to ± 0.001 inches to maintain 99% efficiency. An error of about ± 0.002 inches was found to be acceptable for the ripple magnitude, Δr , while significant errors in the ripple period, λ_B , were permissible.

III. Experimental Results

The device depicted in Fig. 2 was fabricated in 3 sections, one section of 3 ripples, and two containing two ripples each.

In addition to making the fabrication easier, dividing the device up in this manner also allows experimental measurement of the mode content after 2, 3, 4 and 5, as well as the total number of ripples, 7. Two additional tapers were fabricated, one from the existing rectangular to circular waveguide transition output radius of 0.184 inches to the converter input radius of 0.462 inches, and one from the converter output radius of 0.462 inches to the proposed transmission line radius of 0.875 inches.

As we discussed earlier the mode converter is tested reciprocally as a $TE_{11} - TE_{12}$ mode converter. A block diagram of the experimental set-up used on the antenna range is shown in Fig. 5. A circular waveguide taper connects the rectangular to circular transition output diameter to the input diameter of the converter. The TE_{11} mode undergoes some mode conversion in this taper, and the TE_{11} plus the spurious TE_{1n} and TM_{1n} modes then enter the rippled sections. The mode most strongly coupled in this taper, the TM_{11} mode, was measured to be at level of approximately 22.0 dB below the TE_{11} power at the first taper output. The method of determining mode content via pattern measurement is discussed in Appendix A. This slightly impure signal then passes through the mode converter, and through an up taper (which also causes a slight amount of additional mode conversion) to the final diameter of 1.75 inches. By inserting a variable number of rippled sections between the two tapers, and measuring the far field pattern of the taper-ripple-taper chain, the mode content after 0, 2, 3, 4, 5 and 7 ripples was determined.

The patterns measured when no sections are inserted between the two tapers are shown in Fig. 6, along with the ideal TE_{11} patterns which would be measured if there were no mode conversion in the two tapers. Figure 7 shows the measured patterns when the entire converter (7 sections) is inserted between the tapers. The theoretical TE_{12} patterns, which would be measured if the tapers caused no mode conversion and the mode converter were perfect, are also plotted in Fig. 7. The intermediate patterns measured when 3 sections were inserted between the tapers are shown in Fig. 8. Measurements were also taken for 2, 4 and 5 ripple configurations, at the design frequency of 34.5 GHz.

By using the methods described in Appendix A, the mode composition at the output of the final taper was estimated for each of the configurations mentioned above. In particular, estimates for the 0, 3, and 7 ripple cases are summarized in Table 1. The estimated mode content from the measurements is also plotted on Fig. 3 for comparison with theory.

A final set of measurements was made to experimentally determine the conversion efficiency vs. frequency characteristics of the mode converter. For these measurements the

total converter was inserted between the tapers, and pattern measurements were taken for frequencies of 34.0, 34.2, 34.4, 34.5, 34.6, and 34.8 GHz. The mode composition was then determined and the results for the measured efficiency vs. frequency are plotted in Fig. 4 for comparison with theory.

IV. Discussion

In comparing the experimental and theoretical results for the patterns of the tapers connected together with no rippled sections between them, we see that excellent agreement is found between the theoretical TE_{11} H plane pattern and the measured H -plane pattern. The agreement in the E plane is not as good. This is expected since the taper system produces spurious TM_{11} and TM_{12} modes whose effects are only seen in the E plane. Figure 8 again shows good agreement in both planes, with more error in the E plane. This can be explained by considering the taper effects again. The short 0.368–0.924 inch taper generates some spurious TM_{11} power which enters the mode converter. Computer simulations show that spurious TM_{11} power entering the device will be converted primarily into TM modes at the output of the device. Once again, the effects of these modes are only visible in the E plane. The slight asymmetries which are detectable in the measured patterns are probably due to slight misalignments between the taper and mode converter sections.

When the theoretical mode content along the converter and that derived by pattern measurement after 0, 2, 3, 4, 5 and 7 ripples are compared in Fig. 3, excellent agreement is found for the TE_{11} and TE_{12} modes. However, poor agreement is found for the TM_{11} and TM_{12} modes. As was mentioned earlier a spurious TM_{11} signal at a level of approximately –22 dBc was found at the output of the first taper. This signal then enters the rippled sections where it is converted primarily into TM modes. These spurious effects overwhelm the calculated TM_{11} and TM_{12} conversion effects for a pure TE_{11}

mode input, but have little effect on the dominant TE_{11} – TE_{12} interaction. This explains the good agreement seen for the TE_{11} and TE_{12} modes and poor agreement for the TM modes. Despite the spurious effects an overall efficiency over 99.5% was measured for the device (See Table 1), which is also in good agreement with the calculated value of 99.77%.

For the bandwidth results shown in Fig. 4 the best agreement between theory and measurement is found for the points within 100 MHz of the design frequency. For frequencies further removed from 34.5 GHz some of the disagreement may be attributed to errors in the theoretical calculations. For all frequencies, four forward traveling modes were used to model the interaction in the ripples. For the higher frequencies shown in the figure this is probably not sufficient since the TE_{13} mode may also propagate. For the lower frequencies the TM_{12} mode becomes trapped in between the ripples, and reflected waves which are not included in the computer model become important.

V. Conclusions

In conclusion, theoretical and experimental results for a prototype TE_{12} – TE_{11} mode converter have been presented. Good agreement between theory and experiment was found in most cases, and reasonable explanations have been given for the instances where some disagreement has been found. The computer code used to generate the theoretical results presented in this report may now be used with confidence in the design of multimode tapers, mode converters and other devices. The pattern measurement technique for mode identification has also been proved, and may be used to characterize other components. Future work will include an upgrade of the computer code allowing the inclusion of reflected modes. This places confidence in the design of a possible future mode converter which may be placed in the 1.75 inch diameter transmission line for use with a system configured as shown in Fig. 1 [b].

Acknowledgments

The author would like to thank H. Reilly who supervised the fabrication of the tapers and mode converter sections, and S. Stride who assisted in the far field pattern measurements.

References

1. Bhanji, A., Hoppe, D., Hartop, R., Stone, E., Imbriale, W., Stone, D., and Caplan, M., High power Ka-band transmitter for planetary radar and spacecraft uplink, *TDA Progress Report 42-78* (April-June 1984), Jet Propulsion Laboratory, Pasadena, CA, pp. 24-48.
2. Hoppe, D., Imbriale, W., and Bhanji, A. The effects of mode impurity on Ka-band system performance, *TDA Progress Report 42-80* (Oct.-Dec. 1984), Jet-Propulsion Laboratory, Pasadena, CA, pp. 12-23.
3. Hoppe, D., Propagation and radiation characteristics of a multimode corrugated waveguide feedhorn, *TDA Progress Report 42-82*, Jet Propulsion Laboratory, Pasadena, CA.
4. Levine, J., Rippled wall mode converters for circular waveguide, *International Journal of Infrared and Millimeter Waves*, Vol. 5, No. 7, pp. 937-952, 1984.
5. Moeller, C., Mode converters used in doublet III ECH microwave system, *Int. J. Elect.*, Vol. 53, No. 6, pp. 573-585, 1982.

Table 1. Measured TE_{11} - TE_{12} mode converter performance. Mode composition after N mode converter sections.

Mode	Measured			Theory
	$N = 0$	$N = 3$	$N = 7$	$N = 7$
TE_{11} , %	98.23	59.09	*	0.005
TM_{11} , %	1.66	2.45	0.19	0.17
TE_{12} , %	*	37.99	99.57	99.77
TM_{12} , %	0.0025	0.48	0.21	0.05
Others, %	*	*	0.025	0

* No power measured down to system sensitivity.

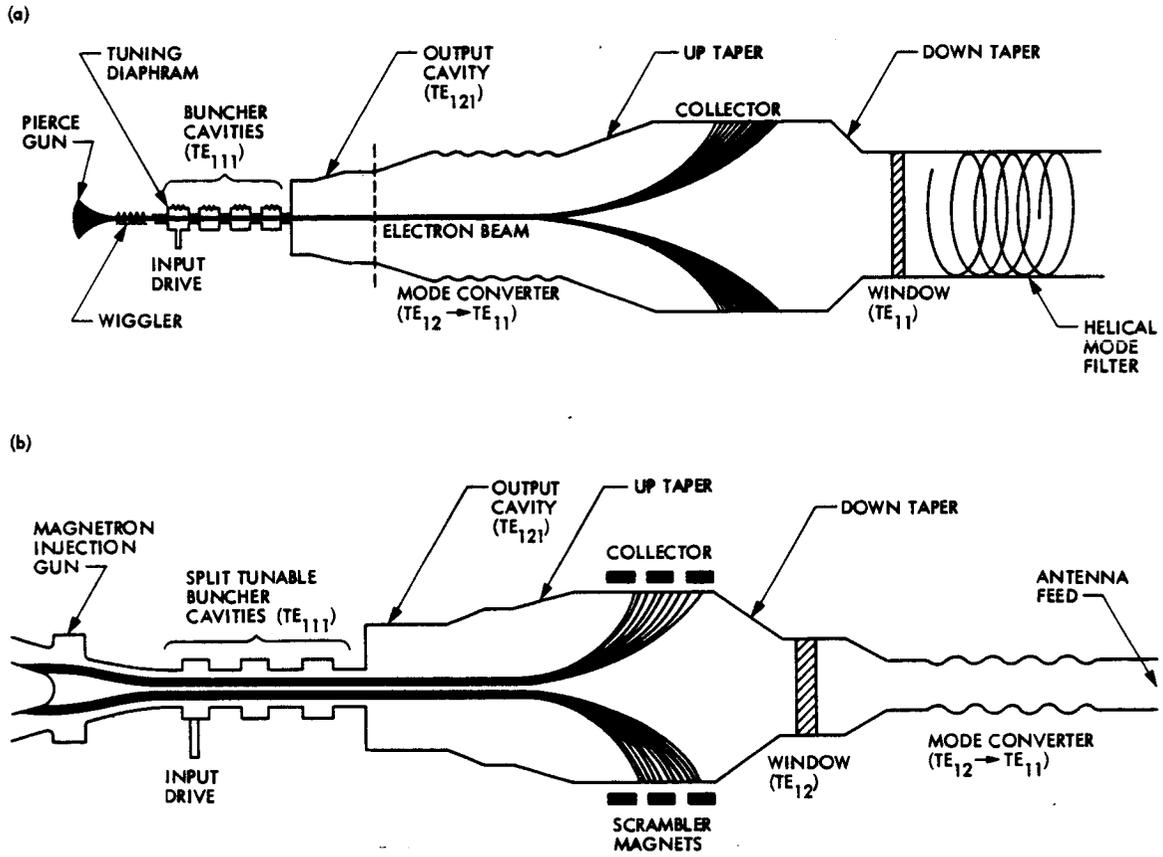


Fig. 1. Two possible gyrokystron configurations: (a) phase 1, (b) phase 2

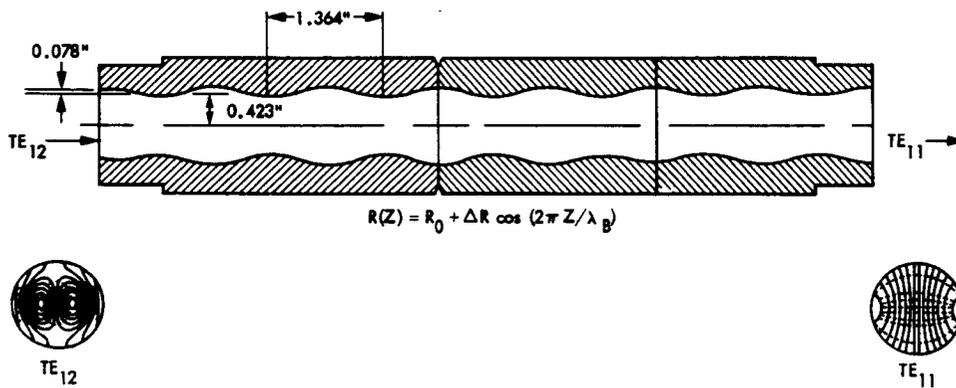


Fig. 2. TE_{12} - TE_{11} mode converter

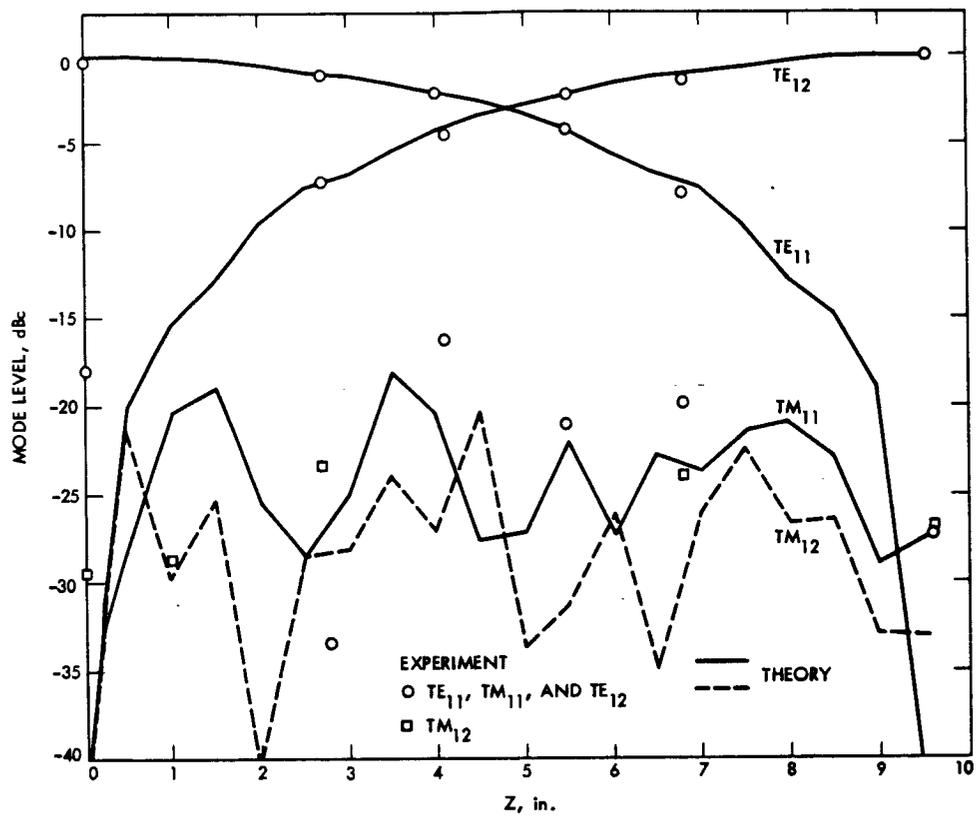


Fig. 3. Mode composition as a function of longitudinal coordinate for the $TE_{11}-TE_{12}$ mode converter

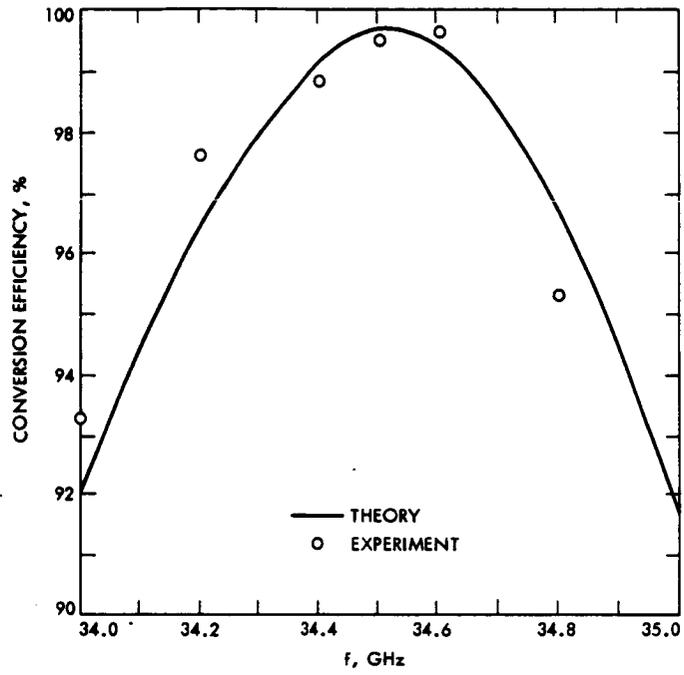


Fig. 4. $TE_{11} - TE_{12}$ mode conversion efficiency vs frequency

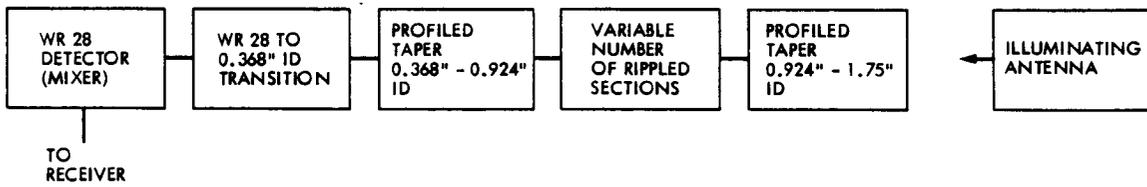


Fig. 5. Antenna range test set up

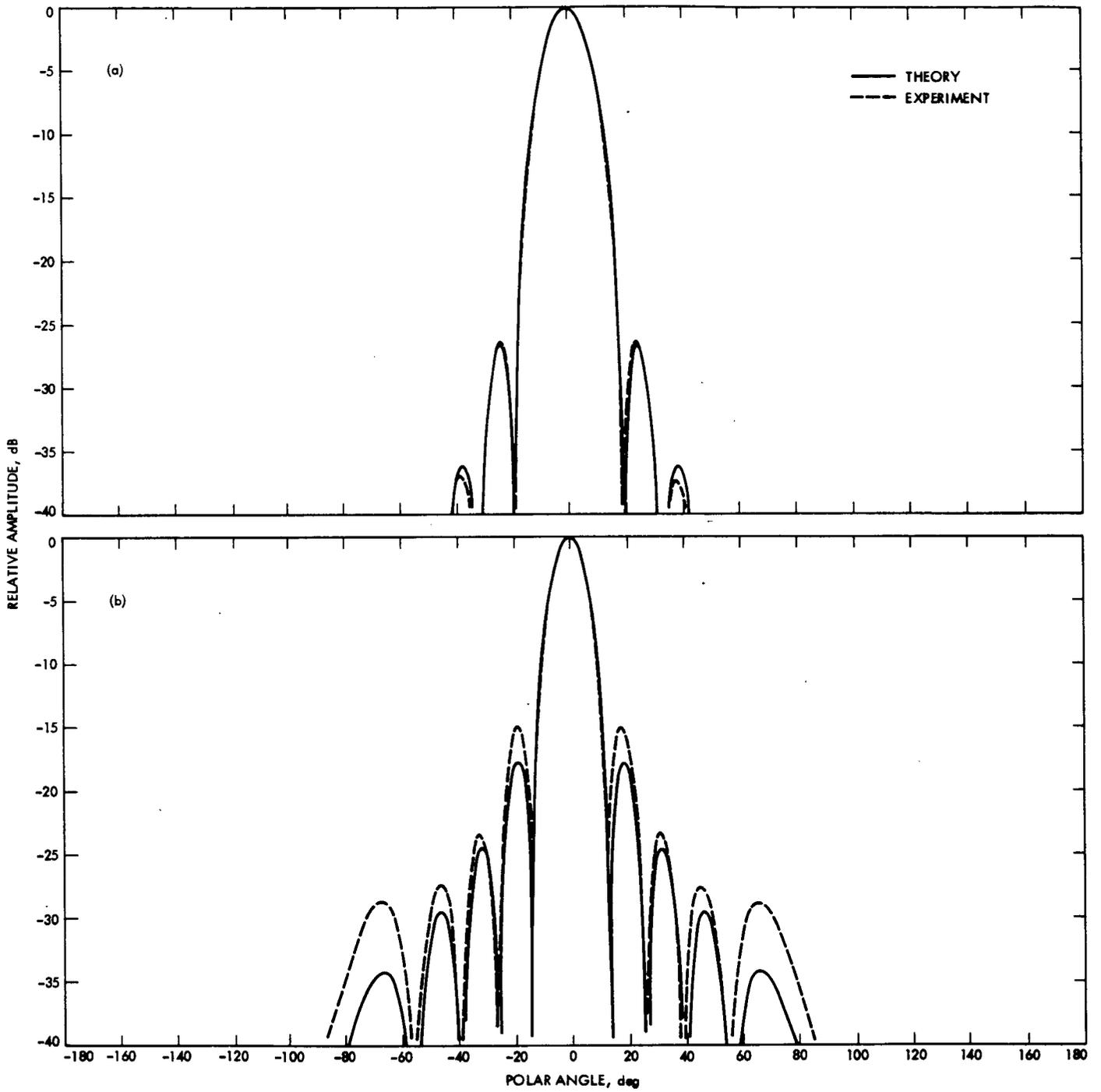


Fig. 6. Taper radiation patterns: (a) H plane patterns, (b) E plane patterns

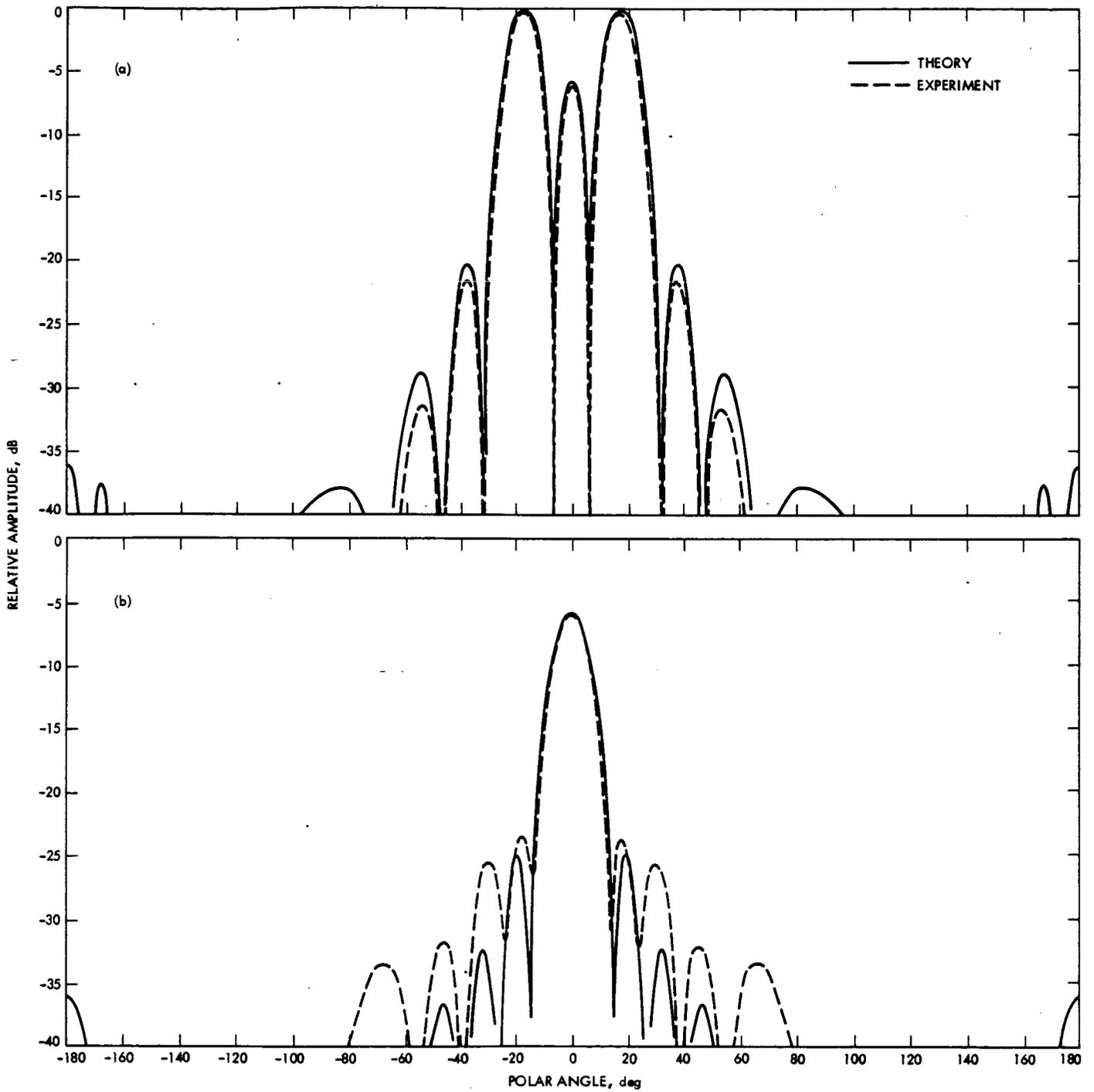


Fig. 7. Total mode converter radiation patterns: (a) H plane patterns, (b) E plane patterns

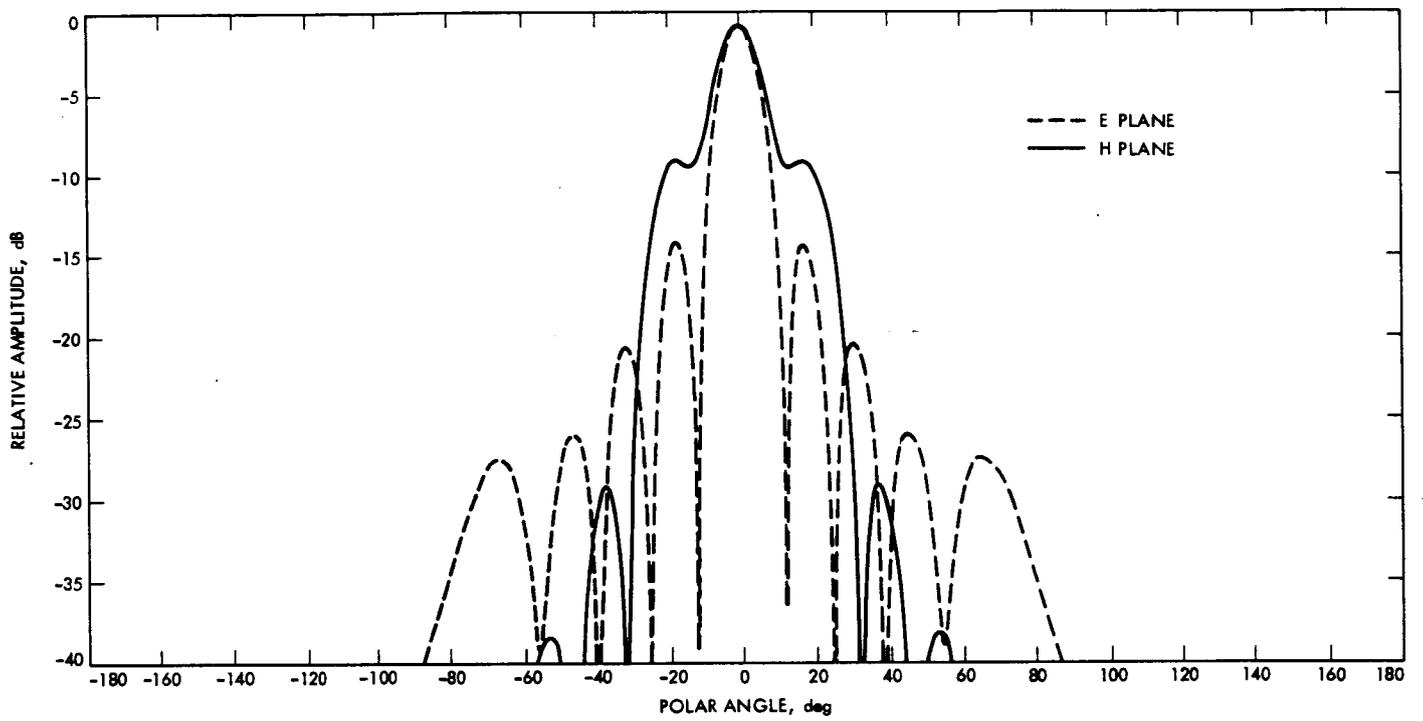


Fig. 8. Three ripple mode converter. Measured radiation patterns.

Appendix A

Mode Identification Using Far Field Radiation Pattern Measurements

In this appendix a method of mode identification in a multimode circular waveguide is explained briefly.¹ The far field radiation pattern for an arbitrary combination of circular waveguide modes assuming small reflection from the aperture is given by (Refs. A-1, A-2):

$$\begin{aligned}
 E(R, \theta, \phi) = & \frac{k \exp(-jkR) (1 + \cos \theta)}{2R} a^2 \\
 & \times \sum_{m,n}^{M,N} \left[\left(C_{mn} (P_{mn})^{1/2} \exp(j\phi_{mn}) \frac{J_m(ka \sin \theta)}{ka \sin \theta} \right. \right. \\
 & \left. \left. - \tilde{C}_{mn} (\tilde{P}_{mn})^{1/2} \exp(j\tilde{\phi}_{mn}) ka \sin \theta \frac{J_m(ka \sin \theta)}{X_{mn}^2 - (ka \sin \theta)^2} \right) \sin m \phi a_\theta \right. \\
 & \left. \times C_{mn} (P_{mn})^{1/2} \exp(j\phi_{mn}) \frac{J'_m(ka \sin \theta)}{(X'_{mn})^2 - (ka \sin \theta)^2} \cos m \theta a_\phi \right] \quad (A-1)
 \end{aligned}$$

where

$$k = 2\pi/\lambda_0$$

$$\lambda_0 = \text{Wavelength}$$

$$R = \text{Far field radius}$$

$$a = \text{Waveguide radius}$$

$$P_{mn} = \text{Power carried by the } TE_{mn} \text{ mode}$$

$$\tilde{P}_{mn} = \text{Power carried by the } TM_{mn} \text{ mode}$$

$$\phi_{mn} = \text{Phase of the } TE_{mn} \text{ mode}$$

$$\tilde{\phi}_{mn} = \text{Phase of the } TM_{mn} \text{ mode}$$

$$C_{mn} = \text{Normalization constant for the } TE_{mn} \text{ mode}$$

$$\tilde{C}_{mn} = \text{Normalization constant for the } TM_{mn} \text{ mode}$$

$$X_{mn} = \text{Zero of } J_m(X)$$

$$X'_{mn} = \text{Zero of } J'_m(X)$$

M, N = Indices chosen large enough to include all propagating modes in the waveguide.

Equation (A-1) assumes that only one of the two orthogonally polarized sets of modes, the set where $E_r \propto \sin \phi$, exists in the waveguide. If we further assume only modes with one azimuthal variation exist, $m = 1$, as in the $TE_{12} - TE_{11}$ mode converter, or tapers, we may write the following equations for the form of the radiation in the E and H planes, respectively, as

E plane

$$\begin{aligned}
 E \propto \sum_{n=1}^N \left(C_{1n} (P_{1n})^{1/2} \exp(j\phi_{1n}) \frac{J_1(ka \sin \theta)}{ka \sin \theta} \right. \\
 \left. - \tilde{C}_{1n} (\tilde{P}_{1n})^{1/2} \exp(j\tilde{\phi}_{1n}) \frac{ka \sin \theta J_1(ka \sin \theta)}{X_{1,n}^2 - (ka \sin \theta)^2} \right) \quad (A-2)
 \end{aligned}$$

H plane

$$E \propto \sum_{n=1}^N C_{1n} (P_{1n})^{1/2} \exp(j\phi_{1n}) \frac{J'_1(ka \sin \theta)}{(X'_{1,n})^2 - (ka \sin \theta)^2} \quad (A-3)$$

From Eq. (A-3) we see that only TE modes contribute to the radiation in the H plane. Furthermore, if we examine the radiation in this plane at the point $ka \sin \theta = X'_{1p}$, where $J'_1(X'_{1p}) = 0$, it is found that all TE_{1n} modes have a null at this point, except the mode TE_{1p} where the denominator and numerator in Eq. (A-3) vanish. Thus, by examining the H plane pattern at the points $ka \sin \theta = X'_{1n}$ $n = 1, \dots, N$ the relative levels of all the TE_{1n} modes may be determined, in terms of power, since the normalization constants C_{1n} are known.

Similarly the relative levels of the TM_{1n} modes may be determined by examining the E plane pattern at the points where $ka \sin \theta = X_{1n}$. In this manner all of the TE_{1n} and TM_{1n} modes may be identified. The method may be extended to include the orthogonal set of $m = 1$ modes by including polarization considerations, and the modes with other azimuthal variation, $m = 0, 2, 3, \dots$ may be identified by taking more pattern cuts.²

¹A more detailed discussion can be found in Z. Zhang, M. Thumm, and R. Wilhelm, "Far field radiation patterns from open-ended oversized circular waveguides and identification of multimode outputs of gyrotrons," Institute fur Plasma fur schung, Universtat Stuttgart (internal document).

²Z. Zhang, M. Thumm, and R. Wilhelm, *op. cit.*

References

- A-1. Silver, S., *Microwave Antenna Theory and Design*, Rad. Lab. Ser., Vol. 12. New York: McGraw-Hill, pp. 336-338, 1949.
- A-2. Ludwig, A. C., Radiation pattern synthesis for circular aperature horn antennas, *IEEE Trans. Antennas and Propagation*, Vol. AP-14, pp. 434-440., July 1966.

Propagation and Radiation Characteristics of a Multimode Corrugated Waveguide Feedhorn

D. Hoppe

Radio Frequency and Microwave Subsystems Section

This report describes a prototype of the multimode corrugated feedhorn which will be used in the 400 kW CW Ka-band radar system. A rough design is done using coupled mode theory and standard corrugated waveguide modes. A more exact analysis using mode matching techniques is then used which takes into account the effect of a finite number of corrugations per wavelength and determines the modes which are reflected from the device. A prototype feedhorn has been constructed and measured. These experimental results are then compared to the theoretical predictions which agree satisfactorily closely.

I. Introduction

A previous report described a conceptual design for a 400 kW CW Ka-band radar transmitter (Ref. 1). In order to transport this large amount of CW millimeter wave power, a multimode transmission line must be used. By using optimally designed waveguide tapers, mode converters, and monitoring devices nearly all of the microwave power at the end of the transmission line will be contained in a rotating TE_{11} circular waveguide mode. The chosen diameter of the multimode waveguide system is essentially the correct size for optimally illuminating the subreflector, but the beamwidth in the E and H planes should be equalized for best antenna efficiency (Ref. 2). Equal E and H plane beamwidths may be obtained by transforming the TE_{11} mode into the balanced HE_{11} mode

in multimode corrugated waveguide. This report describes the theoretical and experimental results for a multimode corrugated waveguide section that accomplishes the above transformation.

II. Theory

Two types of horn antenna possessing equal E and H plane patterns are commonly used. The first is the dual mode horn (Ref. 3), and the second is the corrugated horn (Ref. 4).

In the dual mode horn a dominant mode circular waveguide is connected to another guide of slightly larger diameter, where modes up to TM_{11} may propagate, via a step transition.

The step size is chosen to generate the precise amount of TM_{11} mode from the TE_{11} mode so that when the two modes travel through the flared horn section which follows, the E and H plane patterns are equalized. The bandwidth of this horn is limited since the two modes must arrive at the horn aperture in phase, and the two modes have phase velocities which vary differently with frequency.

In the corrugated horn the single mode smooth wall waveguide is connected to a corrugated waveguide which supports only the HE_{11} mode. Some matching between the waveguides is provided by gradually changing from $\lambda/2$ slot depth to $\lambda/4$ slot depth in a short transition region. Throughout the transition region only the HE_{11} corrugated waveguide mode may propagate, and the E and H plane radiation patterns of this mode become nearly equal when the balanced condition is reached (slot depth $\approx \lambda/4$). The bandwidth of this horn is larger than that of the dual mode horn since the transverse electric field pattern and hence the radiation pattern of the HE_{11} mode are relatively insensitive to small changes in slot depth around the balanced condition (slot depth $\approx \lambda/4$). After the HE_{11} mode is established in the single mode corrugated waveguide, the guide is gradually flared, without changing the slot depth, to the required aperture size.

Special problems arise when multimode smooth and corrugated waveguides are used. In the specific system considered in Ref. 1 a 1.75" I.D. circular waveguide is used at 34.0 GHz allowing over 100 waveguide modes to propagate. The radiation pattern of the TE_{11} mode from the open ended multimode waveguide may be improved by either abruptly converting some of the energy to the TM_{11} mode, as in the dual mode horn, or gradually changing the TE_{11} mode into the balanced HE_{11} mode as in the corrugated horn.

Any attempts to mimic the performance of the step in the dual mode horn in a multimode waveguide are unsuccessful, since a single step transition, slot, or iris generates a large number of modes other than TM_{11} . These modes ruin the pattern symmetry of the TE_{11} - TM_{11} combination.

The second alternative is to gradually change the TE_{11} mode into the HE_{11} mode using a multimode corrugated waveguide in which the slot depths change from $\approx \lambda/2$ to $\approx \lambda/4$. A cut-away view of such a device is shown in Fig. 1. Under multimode conditions new problems arise in making this transition. A 1.75" diameter corrugated waveguide operated at 34.0 GHz also supports many propagating modes, even when the slot depth is constant. When the depth is changed along the guide, coupling occurs between the various corrugated waveguide modes (Ref. 5), in much the same way as radius changes in smooth multimode waveguide cause mode conversion.

The smooth walled circular waveguide may be considered to be a special case of corrugated waveguide for which the slot depth is zero or $\lambda/2$ at the operating frequency. The propagation in a corrugated waveguide of varying or constant slot depth, D , may be written in terms of coupled wave equations (Ref. 6):

$$\frac{d}{dz} A_i(z) = -j\beta_i(D) A_i(z) - \sum_{\substack{j=1 \\ j \neq i}}^N C_{ij}(D) \frac{dD}{dz} A_j(z) \quad i = 1, \dots, N \quad (1)$$

In this equation $A_i(z)$ is the amplitude of the i th corrugated waveguide mode and $\beta_i(D)$ is the phase constant of the i th mode, which is a function of the slot depth at the particular value of z . If the slot depth is changing along the guide ($dD/dz \neq 0$), then coupling occurs between the possible corrugated waveguide modes. The magnitude of the coupling between modes i and j is represented by $C_{ij}(D)$, and it varies approximately as $1/[\beta_i(D) - \beta_j(D)]$, becoming larger for modes with more equal phase velocities. The quantity $\beta(D)$ is found by solving a transcendental equation whose solutions depend on the guide slot depth. For slot depth = 0 or $\lambda/2$, the solutions reduce to those corresponding to TE and TM modes in the smooth walled waveguide. For depths of approximately $\lambda/4$ and $3\lambda/4$ the solutions are those corresponding to the balanced HE and EH modes of corrugated waveguide. The root for the TE_{11} mode (1.841), $D = \lambda/2$, changes to that of the balanced HE_{11} mode (≈ 2.401) as D decreases to $\lambda/4$. If the coupling term in Eq. (1) is neglected, then we find that the TE_{11} to balanced HE_{11} mode conversion is perfect (100%). This is the case when only one mode propagates in the guide, as in the transition section of the standard corrugated horn.

In the case of a multimode corrugated waveguide where D changes from $\lambda/2$ to $\lambda/4$, the TE_{11} - HE_{11} branch is only coupled significantly to the TM_{11} to EH_{12} branch, which has the closest phase constant. In this case Eq. (1) may be rewritten:

$$\begin{aligned} \frac{d}{dz} A_1(z) &= -j\beta_1(D) A_1(z) - C_{12}(D) \frac{dD}{dz} A_2(z) \\ \frac{d}{dz} A_2(z) &= -j\beta_2(D) A_2(z) - C_{21}(D) \frac{dD}{dz} A_1(z) \end{aligned} \quad (2)$$

Here,

$$A_1(0) = TE_{11} \text{ input power}$$

$$A_2(0) = TM_{11} \text{ input power} = 0$$

$$A_1(\ell) = HE_{11} \text{ output power}$$

$$A_2(\ell) = EH_{12} \text{ output power}$$

$$\ell = \text{transition section length}$$

The design objective is to determine a slot depth profile where D varies from $\lambda/2$ to approximately $\lambda/4$ which maximizes $A_1(\ell)$, and minimizes $A_2(\ell)$ when Eqs. (2) are solved numerically. Several profiles for the slot depth variation were considered, including a linear profile, and one which provides constant coupling along the transition section, making

$$C_{12}(D) \frac{dD}{dz} = K \quad (3)$$

where K is a constant (Ref. 6). Of the profiles examined, this profile gave maximum conversion efficiency in a minimum length ℓ . For this profile the slot depth changes most gradually near depths of $\lambda/2$, where the phase constants of the two branches are closest, and more rapidly as D approaches $\lambda/4$. This profile was chosen for the actual device which was built and tested.

Certain assumptions are made in the above analysis. In particular, an ideal corrugated waveguide is modeled where an infinite number of slots per wavelength are assumed. Reflections are ignored, and only the main coupling interaction is considered. A more exact analysis of the problem which eliminates all of these faults can be undertaken. In this method the corrugated waveguide is modeled as a large number of straight waveguide sections of different radii which are connected in series. Electric fields are described as sums of smooth walled circular waveguide modes, and are matched at each of the discontinuities. The result is a scattering matrix which relates the reflected circular waveguide modes at the input to the device and the output modes, to the incident modes at the input end. The theory of this approach is given in detail by James (Ref. 7), and a computer code to carry out the computations has been written by the author¹ (see also Ref. 8, section B). The scattering matrix approach may be summarized by two matrix equations:

$$\begin{aligned} b_1 &= [S_{11}] a_1 \\ b_2 &= [S_{21}] a_1 \end{aligned} \quad (4)$$

¹See D. J. Hoppe, "Scattering Matrix Program for Circular Waveguide Junctions," Interoffice Memorandum No. 3335-84-071 (internal document), Dec. 5, 1984, Jet Propulsion Laboratory, Pasadena, CA.

Here a_1 contains the input mode amplitudes, b_1 the reflected modes, and b_2 the output modes. The matrices $[S_{11}]$ and $[S_{21}]$ are scattering matrices which are determined by the computer code. Their entries are a function of frequency and device geometry. This computer program was used extensively in another development (Ref. 8), which calculated the theoretical radiation patterns for the feed under different excitation conditions.

A device of the type described above was designed using the simple coupled mode theory. The more exact scattering matrix theory was then used to predict the radiation patterns and reflection characteristics of the corrugated waveguide section. The part was then built and measured on the antenna range. These theoretical and experimental results are described in the next section.

III. Calculated and Experimental Results

The experimental model of the corrugated feed consisted of 92 slots cut in the wall of a 1.75 inch I.D. circular waveguide. The width of the slots and lands was chosen to be 0.035 inch. These parameters were chosen for the test device for several reasons. First, Eq. (2) was solved for a device of various lengths using the constant coupling slot depth profile. From this rough analysis it was found that feed lengths in excess of 15λ gave an HE_{11} mode purity level above 95%. Next a very conservative number of slots per wavelength, 5, was chosen for the test device. The final consideration was the machining complexity. It was decided to keep the number of slots under 100, and 92 were chosen which corresponded to an HE_{11} mode purity of 96.5% from the simple analysis. The slot depth profile was determined by the methods described in the previous section, and each slot depth was specified individually. The final design parameters were then used as input data for the scattering matrix program which produced the final theoretical radiation patterns. The detailed calculation of the scattering matrix for the device involved the cascading of 370 40-by-40 scattering matrices representing the straight sections and waveguide junctions making up the device.

The scattering matrix method calculates the matrix for each subsection of the total device (i.e., after 1 slot, 2 slots, . . . , up to 92 slots). From this we may determine the aperture modes that would be present if the device were terminated after any number of slots. The results of this calculation for a frequency of 34 GHz are shown in Fig. 2. The figure shows the increasing level of TM_{11} and decreasing level of TE_{11} as more slots are added, as well as a few of the other modes which are present. The final aperture modes and reflected modes are shown for the 92 slot device in Table 1. The aperture modes were then used to determine the E and H plane far field patterns of the feed. These results are shown in Figs. 3 and 4.

The phase center of the feed was calculated as being exactly in the aperture plane. The efficiency of the 64-meter antenna in its present hyperboloid-paraboloid configuration, neglecting surface tolerance errors, was calculated using the theoretical patterns of Figs. 3 and 4. This efficiency was calculated to be 80.8% using the equivalent paraboloid method, and 79.5% using the GTD program (Ref. 8). Some theoretical results for the bandwidth were also determined. The calculated feed patterns for frequencies of 33.5 and 34.5 GHz are shown in Figs. 5 and 6, respectively. The theoretical antenna efficiency at 33.5 and 34.5 GHz was determined to be 81.0% and 80.4% respectively, both numbers being calculated using the equivalent paraboloid method.

A block diagram of the experimental test set up used on the mesa is shown in Fig. 7. The transition from single mode (0.368 I.D.) to multimode (1.75 I.D.) waveguide is a profiled circular waveguide taper. An ideal taper would produce a perfectly pure TE_{11} mode at the 1.75 I.D. end. For this particular taper the TM_{11} mode, which is the mode most strongly coupled to the TE_{11} mode in the taper, was measured to be about 29 dB below the TE_{11} mode at the taper output.

The measured patterns for the device are plotted in Figs. 3 and 4 for comparison. Excellent agreement is found in both planes with the best agreement in the H-plane. The slight differences in the E-plane may be accounted for by recalling that the calculated pattern is for a pure TE_{11} mode input, but in the experiment some spurious TM_{11} power is also present. The effects of this slight mode impurity appear primarily in the E plane (see Ref. 8, Figs. 4 and 6). The experimental phase center of the feed was also determined to be in the aperture plane, which is also in agreement with theory.

IV. Conclusions

In conclusion, excellent agreement between the theoretical predictions of the scattering matrix program and experiment were obtained for a multimode corrugated feed section at 34 GHz. Since the program is also capable of analyzing corrugated horns of arbitrary profile and various other circular waveguide devices, it is expected to prove to be a valuable tool for analysis in the future.

Acknowledgments

The author would like to thank H. Reilly who supervised the fabrication of the various waveguide components and S. Stride who assisted in the pattern measurements.

References

1. Bhanji, A., Hoppe, D., Hartop, R., Stone, E., Imbriale, W., Stone, D., and Caplan, M., High power Ka-band transmitter for planetary radar and spacecraft uplink, *TDA Progress Report 42-78* (April-June 1984), Jet Propulsion Laboratory, Pasadena, CA, pp. 24-48.
2. Potter, P. D., and Ludwig, A., Antennas for space communications, *JPL Space Program Summary No. 37-26*, Vol. IV, Jet Propulsion Laboratory, Pasadena, CA, April 30, 1984, pp. 200-208.
3. Potter, P. D., A new horn antenna with suppressed sidelobes and equal beamwidths, *Microwave Journal*, June 1963, pp. 71-78.
4. Clarricoats, P. J., and Saha, P. K., Propagation and radiation behavior of corrugated feeds, Part 1, *Proc. IEE*, September 1971, pp. 1167-1176.
5. Dragone, C., Reflection, transmission and mode conversion in a corrugated feed, *Bell System Tech. J.*, July-August 1977, pp. 835-867.
6. Doane, J. L., Mode converters for generating the HE_{11} (Gaussian-like mode) from TE_{01} in circular waveguide, *Int. J. Elect.*, Vol. 53, December 1982, pp. 573-585.
7. James, G. L., Analysis and design of TE_{11} and HE_{11} , corrugated cylindrical waveguide mode converters, *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, October 1981, pp. 1059-1066.
8. Hoppe, D., Imbriale, W., and Bhanji, A., The effects of mode impurity on Ka-band system performance, *TDA Progress Report 42-80* (October-December 1984), Jet Propulsion Laboratory, Pasadena, CA, pp. 12-23.

Table 1. Aperture and reflected modes for a pure TE_{11} mode incident on the feed

Modes	Percent of Incident TE_{11} Power, %	dB Below Incident TE_{11} Power, dBc
Reflected Modes		
TE_{11}	0.0002	-58
TM_{11}	0.0012	-49
TE_{12}	0.0001	-62
TM_{12}	0.0010	-50
TE_{13}	< 0.0001	-65
TM_{13}	0.0015	-48
TE_{14}	< 0.0001	-65
TM_{14}	0.0054	-43
TE_{15}	0.0001	-62
Aperture Modes		
TE_{11}	81.95	-0.86
TM_{11}	17.76	-7.5
TE_{12}	0.054	-33
TM_{12}	0.010	-40
TE_{13}	0.0013	-49
TM_{13}	0.2003	-27
TE_{14}	0.0003	-55
TM_{14}	0.0090	-40
TE_{15}	0.0008	-51

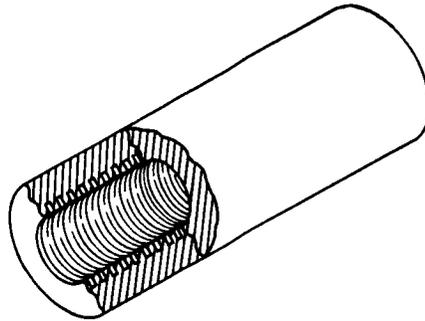
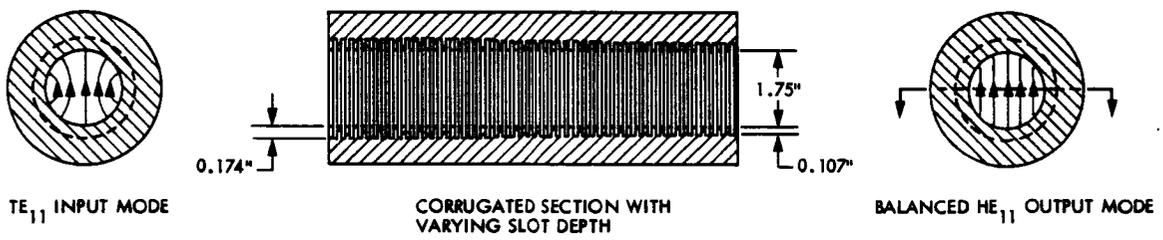


Fig. 1. The antenna feed

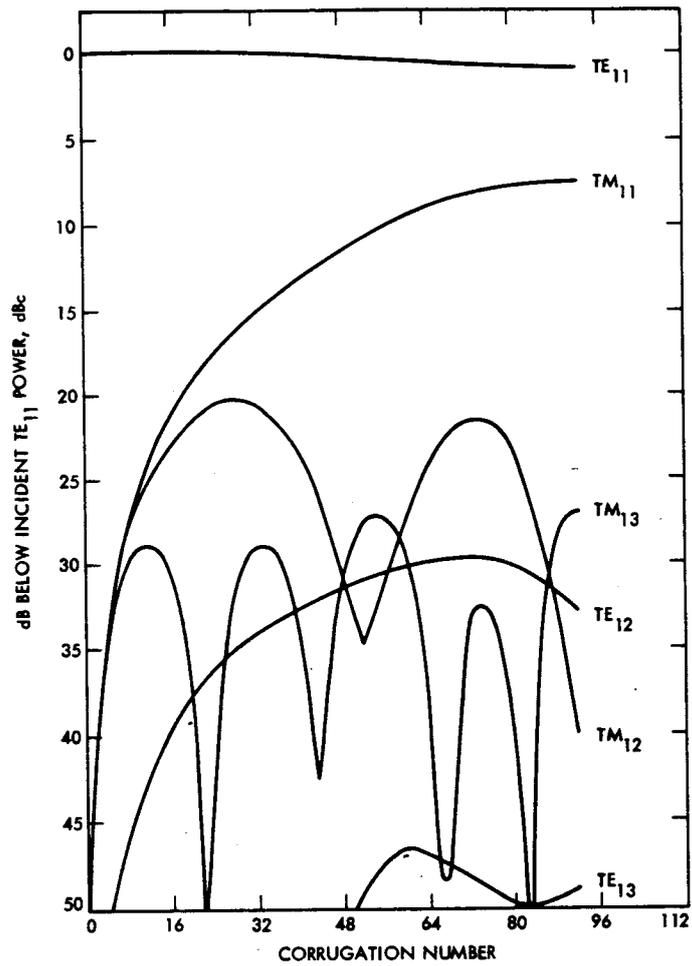


Fig. 2. Mode content vs corrugation number

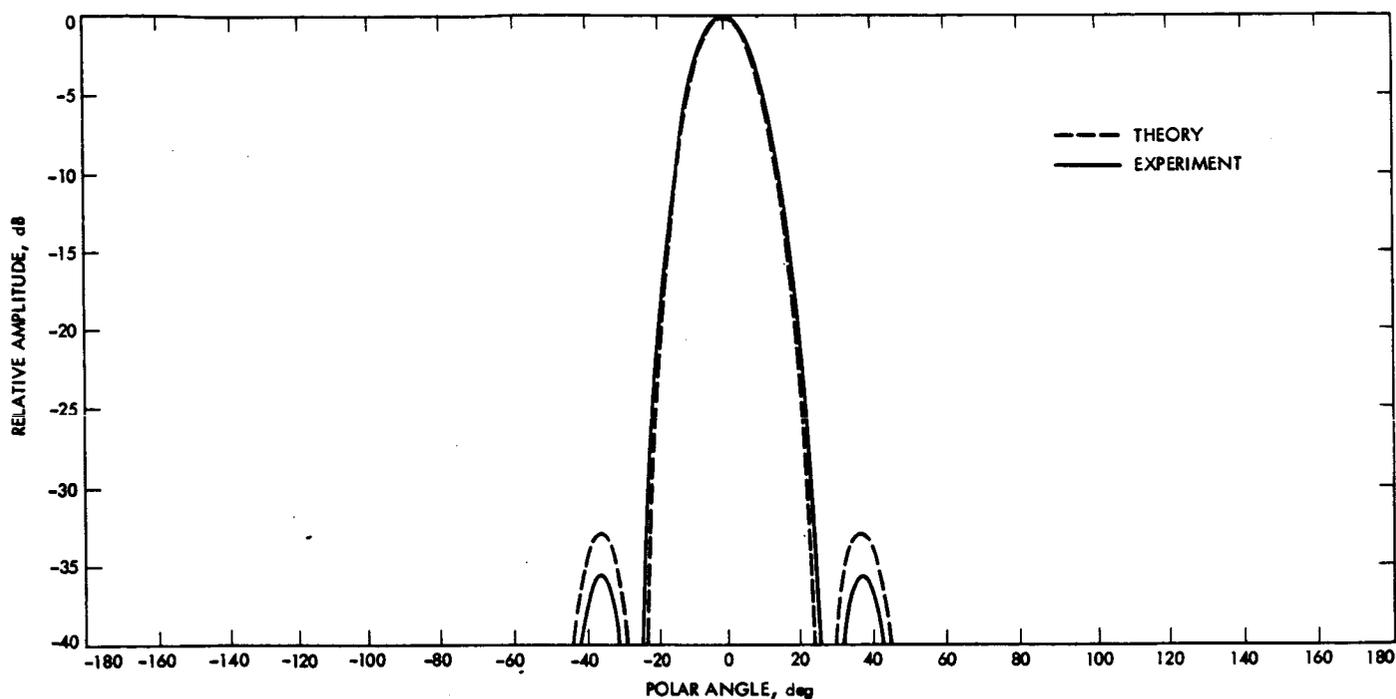


Fig. 3. Theoretical and experimental E-plane patterns at 34.0 GHz

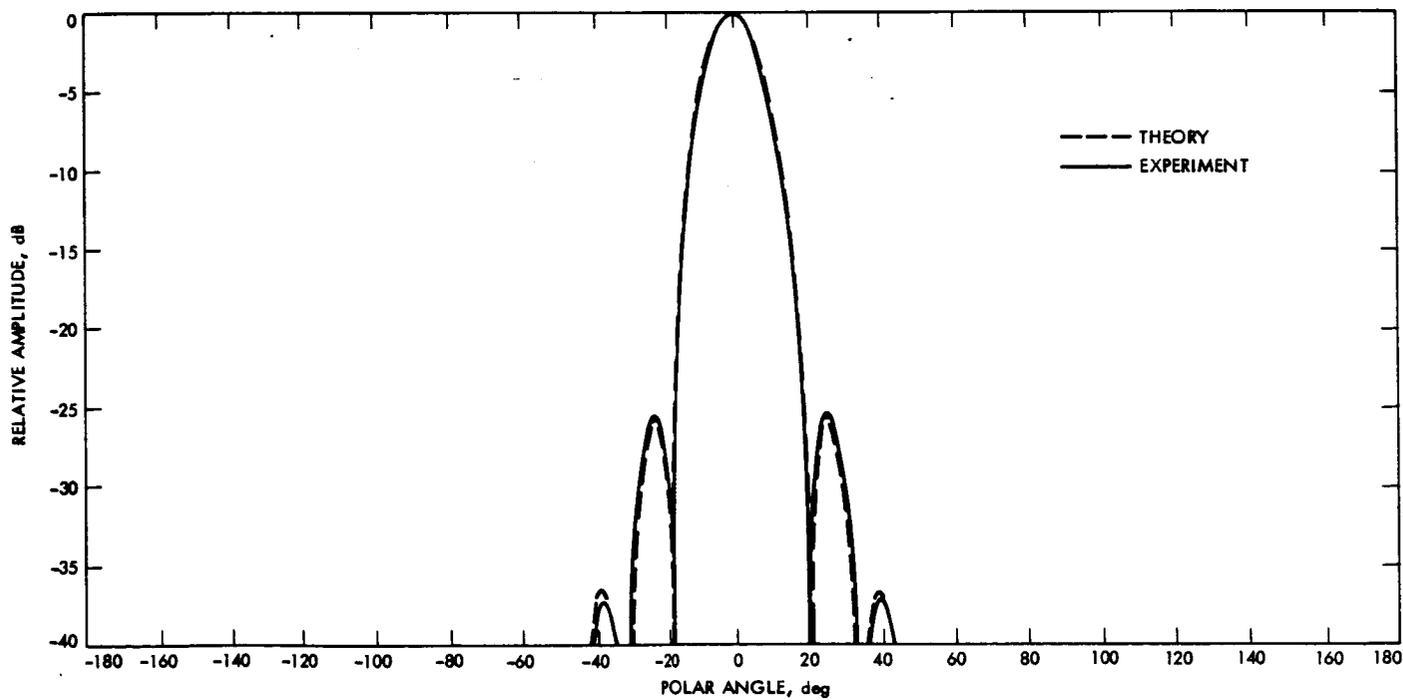


Fig. 4. Theoretical and experimental H-plane patterns at 34.0 GHz

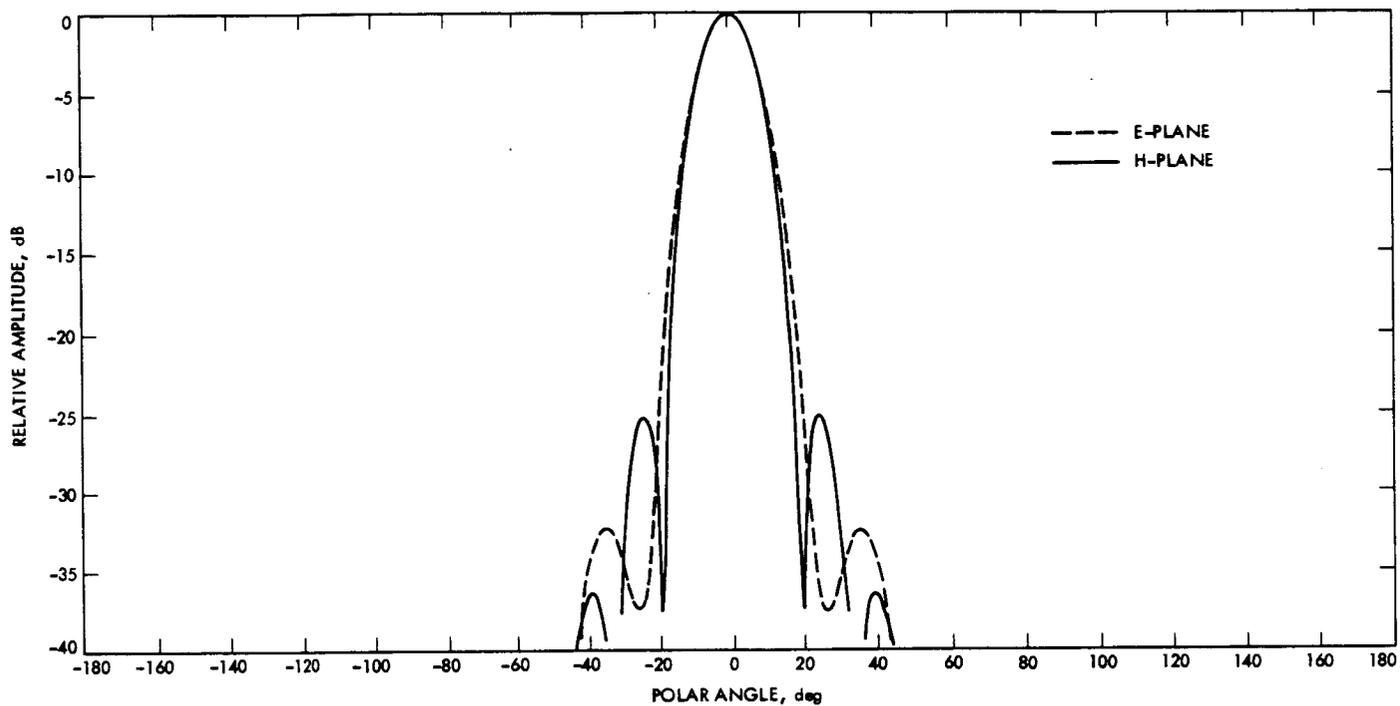


Fig. 5. Theoretical feed patterns at 33.5 GHz

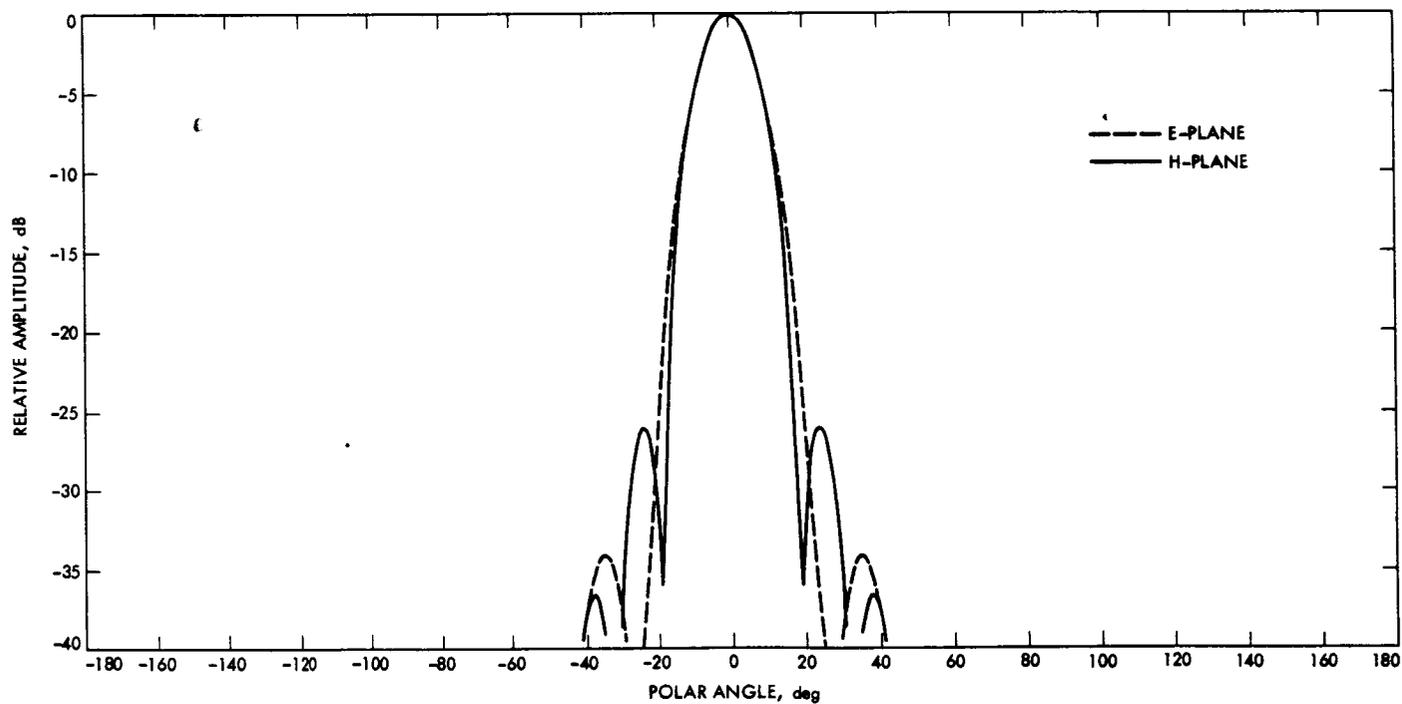


Fig. 6. Theoretical feed patterns at 34.5 GHz

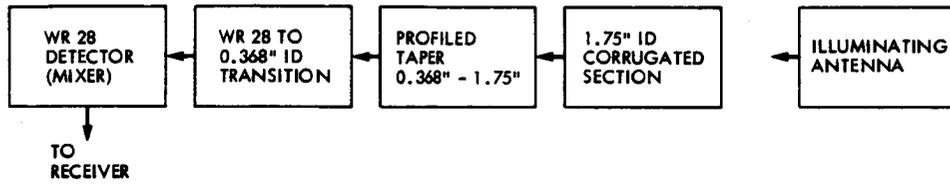


Fig. 7. Antenna range test set-up

Load-Deflection Tests and Computer Analyses of a High-Precision Adhesive-Bonded Antenna Reflector Panel

C. T. Chian and R. Levy

Ground Antennas and Facilities Engineering Section

New adhesive-bonded panels are being investigated as a part of an effort to extend and upgrade the 64-m to a 70-m antenna network. Load-deflection tests were conducted on a sample high-precision adhesive-bonded panel for comparison with design criteria. Two computerized structural models were developed in order to predict the deformation patterns under different types of distributed and concentrated loadings. The main purpose was to obtain empirical stiffness factors for the slit beams and girders in the panel structure. With determination and use of the empirical stiffness factors, there is a good agreement between the theoretically predicted deflections and the test measurements. It was also found that the new bonded panels satisfy the stringent design specifications and surface tolerance bounds.

I. Introduction

New high-precision adhesive-bonded panels are being investigated under the Advanced System Program's research and development effort in order to replace the traditional rivetted panels, reduce fabrication costs, and improve the surface tolerance characteristics.

The new adhesive-bonded panels, if successful, would be used in the current 64-m antenna network rehabilitation and extension project at X-band. The new panels are expected to reduce about 0.5 dB of gain loss at X-band.

II. Panel Description

The test panel is adhesive-bonded and has a solid 0.18-cm (0.070-inch) skin thickness with no skin perforations. It has

circumferential beams at 35.6-cm (14-inch) centers and a radial girder at each of two sides to support the beams as shown in Fig. 1. Beams and girders are zee sections, slit and reinforced (Figs. 2 and 3). The test panel is fabricated under a JPL contract to Toronto Iron Works (TIW) of Sunnyvale, California and a subcontract to COSPAL in Bergamo, Italy.

The purpose of this test was to evaluate the response of the bonded panels for distributed loadings that are representative of environmental wind and gravity loadings, and to determine empirical stiffness and residual deflection factors due to the slitting of girders and beams.

A series of tests were conducted to measure panel deflections due to stimulated wind loadings and specified, concentrated shoe loads up to 135 kg (300 lb). The geometry of the test panels is similar to panel number 8 of the new 34-m

AZ/EL antennas, recently built by TIW for Deep Space Stations 15 and 45. In order to provide a higher structural stiffness, the new test panels were designed with ten back-up circumferential beams, instead of seven, as originally employed for the 34-m antennas.

III. Theoretical Models

Two computerized structural models were developed for simulation of the deflections of the test panel. The first model for theoretical panel deflection analysis is the FORTRAN program PANELDEFL. PANELDEFL was designed to provide analysis for the full set of panel configurations used in a complete antenna. The analysis model is automated using a minimum description of the essential geometry and construction features of each panel of the set. The PANELDEFL program provides a microwave pathlength analysis and summary for the set of all of the panels, as well as for each individual panel. The method of deflection analysis is to integrate the Euler-Bernoulli differential equations for the beams and girders. A more detailed description of the PANELDEFL program can be found in Appendix A.

The second panel structural model was developed by using the NASTRAN finite element program for the tested panel as a check. Since the PANELDEFL program does not use finite element approximations, it is expected to be more accurate than NASTRAN for the present problem. A description of the NASTRAN panel deflection model is given in Appendix B. Comparison of the outputs of the two programs showed the results to be almost identical.

The results of theoretically predicted panel deflections from PANELDEFL were compared with the measured data. This procedure has provided good validation and correlation of the computer models and good estimates of the material elastic properties.

IV. Test Loading Configurations

Five series of tests were conducted on the adhesive-bonded panel, which comprised both the distributed loading and the concentrated loading cases (Table 1). The test configurations were of two kinds:

- A. Panel without center supports
- B. Panel with center supports.

The test bed consists of a steel surface plate and I-beam fixture that provided rigid panel support and measurement points. The whole test bed, together with the sample panel, is shown in Fig. 4. A schematic diagram of the test panel is

given in Fig. 5. A total of 19 dial indicators were used to measure the panel deflections in various locations, which are shown in Fig. 6. The test configurations are described as follows.

A. Panel Without Center Supports

The maximum distributed panel loading was 39 kg/m^2 (8.0 PSF). This distributed load for the sample panel corresponds to a uniform thickness of about 2.54 cm (1.0 inch) of sand. Up to 56 sandbags with predetermined weights were used to represent the distributed loading as shown in Figs. 7 and 8.

The girder bending stress was estimated to be only about 210 kg/cm^2 (3000 psi), which is less than one-tenth of the yield stress. After the distributed load test was completed, the concentrated load test was conducted. A concentrated load of 85.5 kg (190 lb) was used. Steel plates $30 \times 30 \times 2.54 \text{ cm}$ ($12 \times 12 \times 1 \text{ inch}$) were placed on top of a smaller steel plate $10 \times 30 \times 2.54 \text{ cm}$ ($4 \times 12 \times 1 \text{ inch}$) to simulate the concentrated shoe load as shown in Fig. 9.

B. Panel With Center Supports

Both the distributed load test and the concentrated load test were also conducted for this panel test configuration. The distributed load tests were performed first. The maximum distributed loading, with center supports in place, was about 78 kg/m (16.0 PSF) on the gross area, which corresponds to about 5.08 cm (2.0 inches) of sand.

This is a typical estimated wind load that the panel is required to support, allowing for 40% perforation area. This loading occurs when the antenna is tilted to the zenith position under a 160 km/hr (100 mph) wind. Up to 84 sandbags with predetermined weights were evenly distributed on the panel. After the distributed load test was completed, the concentrated load test was then conducted. Concentrated loads of 99, 126, and 135 kg (220, 280, and 300 lb) were used. Again, steel plates of $30 \times 30 \times 2.54 \text{ cm}$ ($12 \times 12 \times 1 \text{ inch}$) were placed on top of a smaller $10 \times 30 \times 2.54 \text{ cm}$ ($4 \times 12 \times 1 \text{ inch}$) to simulate the concentrated shoe load. Each of the $30 \times 30 \times 2.54 \text{ cm}$ steel plates weigh about 18.5 kg (41 lb), while the smaller plate ($10 \times 30 \times 2.54 \text{ cm}$) weighs approximately 6.3 kg (14 lb). Each load increment was intended to have a minimum of 20 minutes duration time. The recorded data reflected this time effect.

V. Comparison Between Theoretical and Test Results

Verification of the two panel structural models was made against the measured data to compare the two different

numerical approaches. Measured deflection data taken from the 19 dial indicators were recorded and compared with the theoretically predicted values.

The results are tabulated in Tables 2 and 3. Table 2 shows comparisons between the two numerical approaches for the following two cases: (1) Panel without center supports, with distributed load at 39 kg/m² (8 PSF); (2) Panel with center supports, with distributed load at 78 kg/m² (16 PSF). The results from the PANELDEFL and NASTRAN programs differed only by 3%.

Empirical stiffness factors for the panel back-up beams and girders were used to account for the slits. These stiffness factors were determined in an iterative process to achieve a good agreement with the measured deflection data. For the sample panel tested, it was found that for Girder Stiffness Factor (GSF) of 1.00 and Beam Stiffness Factor (BSF) of 0.80, the test and theory agrees to within 12%, for most of the data (Table 3). Therefore, moments of inertia for the slit beams, in the current configuration, are computed as 80% of the moments of inertia for the beams without slitting. The rivetted panels had been load-tested previously.

Comparisons of stiffness factors obtained from different tests indicated that the empirical stiffness factors are strongly dependent on the panel configurations.

Table 4 shows the comparison of empirical stiffness factors for panel girders and beams for three tests. A 3-day duration

test was also conducted to study the hysteresis of the panel. A concentrated load of 135 kg (300 lb) was applied at the center of the panel and maintained for three days. Daily readings of the dial indicators were made. However, after the load was removed, no permanent deformation was observed.

VI. Summary and Conclusions

At the specified level of loads, the tested adhesive-bonded panel was found to deflect linearly with the load, for both the distributed and concentrated loading cases as shown in Figs. 10 and 11. Deflections showed insignificant hysteresis when the load is removed, even for the 3-day-long duration test. The adhesive-bonded panel withstood the specified level of wind loads, as well as the concentrated shoe load, without any apparent degradation.

The empirical stiffness factors for the slit beams and girders used in the panel structural models were found to be strongly dependent on the panel configurations. The variation of the panel configurations include whether the panel is rivetted or bonded, and whether the panel skin is solid or perforated. The spacing of beams and the overall geometry of the panel also influence the determination of the empirical stiffness factors. Because of such a wide variation of the panel configurations, and because there is no simple theoretical way to compute or predict the stiffness factors, it is concluded that a test is needed for each type of panel configuration in an antenna set.

Acknowledgments

The authors acknowledge the assistance given by F. Lansing, S. Rocci, and the Ground Antennas and Facilities Engineering Section field engineering technicians during the various steps of this work.

Table 1. Panel load-deflection tests

Test series	Center supports	Loading type	Load level
1	No	Distributed	19.5 kg/m ² (4 PSF); 39 kg/m ² (8 PSF)
2	Yes	Distributed	39 kg/m ² (8 PSF); 78 kg/m ² (16 PSF)
3	Yes	Concentrated	99 kg (220 lb); 135 kg (300 lb)
4	Yes	Concentrated	126 kg (280 lb)
5	No	Concentrated	85.5 kg (190 lb)

Table 2. Comparison of theoretically predicted deflections (Unit: cm (inch), distributed loading)

(a) Panel Without Center Supports, at 39 kg/m² (8 PSF)

Dial Indicator Number / Theory	8	9	10	16	17	21
NASTRAN Model (A)	0.106 (0.0419)	0.102 (0.0404)	0.092 (0.0362)	0.076 (0.0297)	0.096 (0.0379)	0.066 (0.0264)
PANELDEFL Model (B)	0.104 (0.0410)	0.100 (0.0395)	0.089 (0.0352)	0.075 (0.0294)	0.095 (0.0372)	0.065 (0.0259)
Ratio A/B	1.02	1.02	1.03	1.01	1.01	1.01

(b) Panel With Center Supports, at 78 kg/m² (16 PSF)

Dial Indicator Number / Theory	8	9	10	16	17	21
NASTRAN Model (A)	0.0625 (0.0246)	0.0544 (0.0214)	0.0318 (0.0125)	0.0706 (0.0278)	0.0653 (0.0257)	0.0564 (0.0222)
PANELDEFL Model (B)	0.0605 (0.0238)	0.0526 (0.0207)	0.0307 (0.0121)	0.0683 (0.0269)	0.0638 (0.0251)	0.0577 (0.0227)
Ratio A/B	1.03	1.03	1.03	1.03	1.02	0.98

Table 3. Comparison of theory and test data (unit: cm)

(a) Panel Without Center Supports, Distributed Load at 39 kg/m² (8 PSF)

Dial Indicator Number	6	7	8	16	17	20	18	21
Deflection								
Test (A)	0.086	0.099	0.104	0.074	0.104	0.094	0.089	0.061
Theory (B)	0.086	0.097	0.099	0.074	0.092	0.099	0.086	0.064
Ratio A/B	1.00	1.02	1.05	1.00	1.13	0.95	1.03	0.95

(b) Panel With Center Supports, Distributed Load at 78 kg/m² (16 PSF)

Dial Indicator Number	6	7	8	16	17	20	18	21
Deflection								
Test (A)	0.038	0.056	0.061	0.038	0.076	0.066	0.074	0.058
Theory (B)	0.030	0.053	0.061	0.043	0.064	0.061	0.061	0.056
Ratio A/B	1.27	1.06	1.00	0.88	1.19	1.08	1.21	1.03

Table 4. Effective stiffness factor for girders and beams

Test	(7/84, Panel Numbers 3 and 9)			(10/84, Modif. Panel Number 8)		(4/85, Modif. Panel Number 8)	
	Goldstone test on rivetted panel			COSPAL test on bonded panel		Goldstone test on bonded panel	
Center support	No	No	Yes	No	No	No	Yes
Panel skin	Solid	Perf.	Perf.	Solid	Solid	Solid	Solid
No. of beams				10	7	10	10
GFACT ^a	0.60	0.50	0.48	0.87	0.74	1.00	1.00
BFACT ^b	0.75	0.75	0.68	0.90	0.95	0.80	0.80

^aGFACT = Effective stiffness factor for girders.

^bBFACT = Effective stiffness factor for beams.

ORIGINAL PAGE IS
OF POOR QUALITY

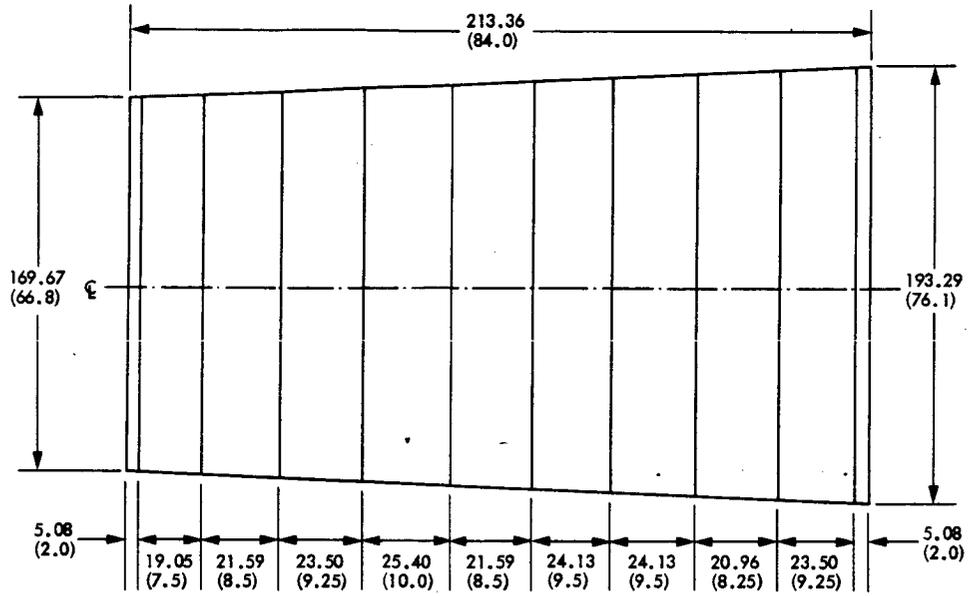


Fig. 1. Panel dimensions and back-up beam spacings [unit in cm (Inch)]

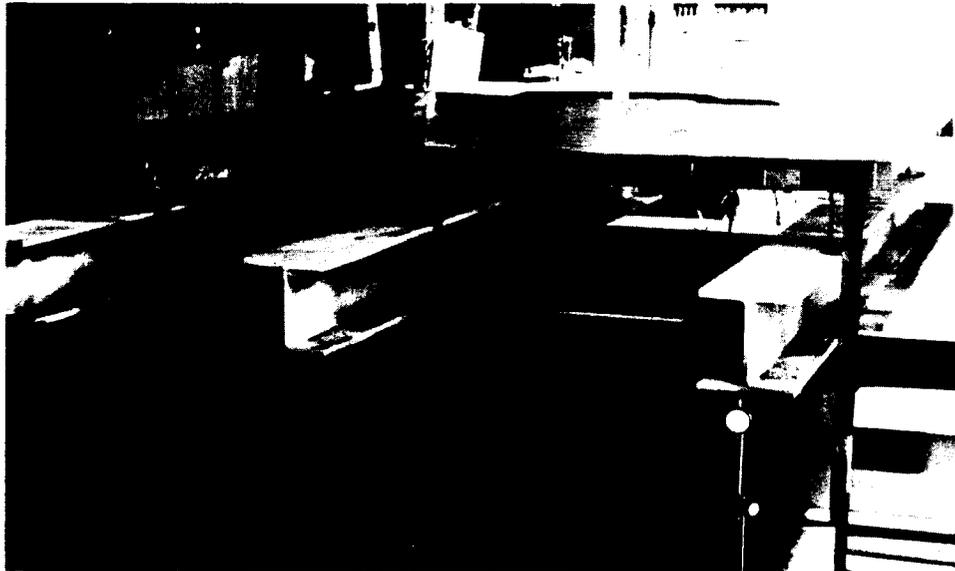


Fig. 2. Test panel showing slit girder

ORIGINAL PAGE IS
OF POOR QUALITY

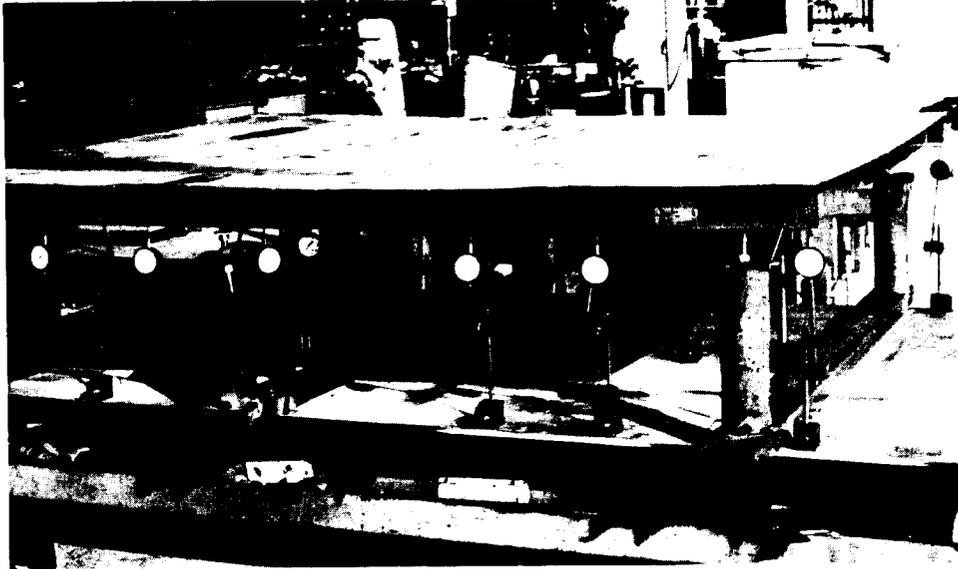


Fig. 3. Test panel with slit beams and girders

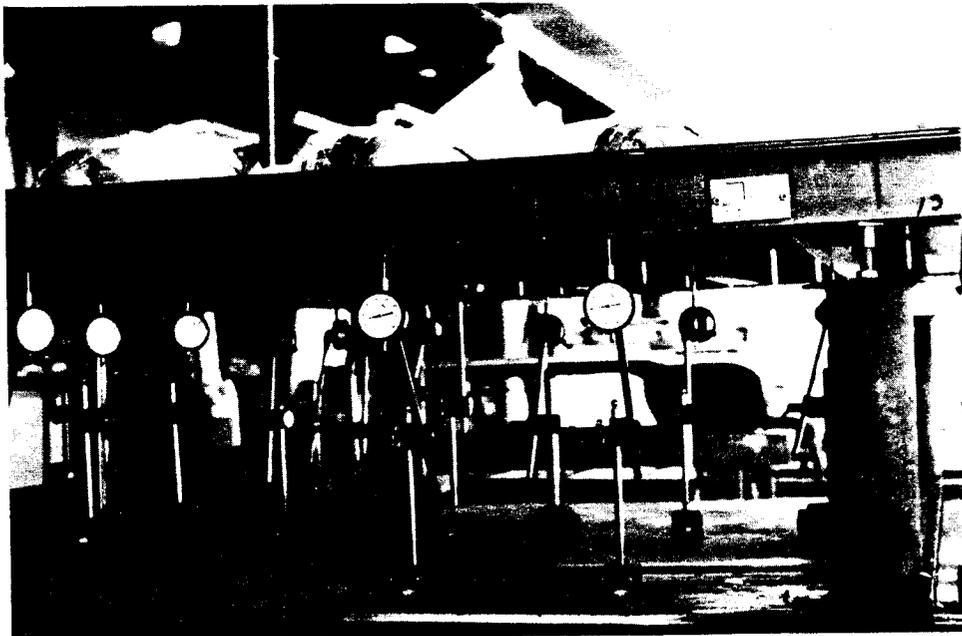


Fig. 4. Sample panel with dial indicators

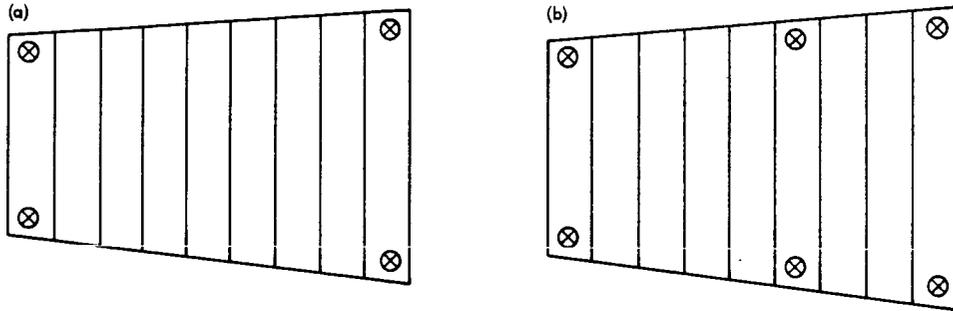


Fig. 5. Panel loading configurations: (a) without center supports; (b) with center supports

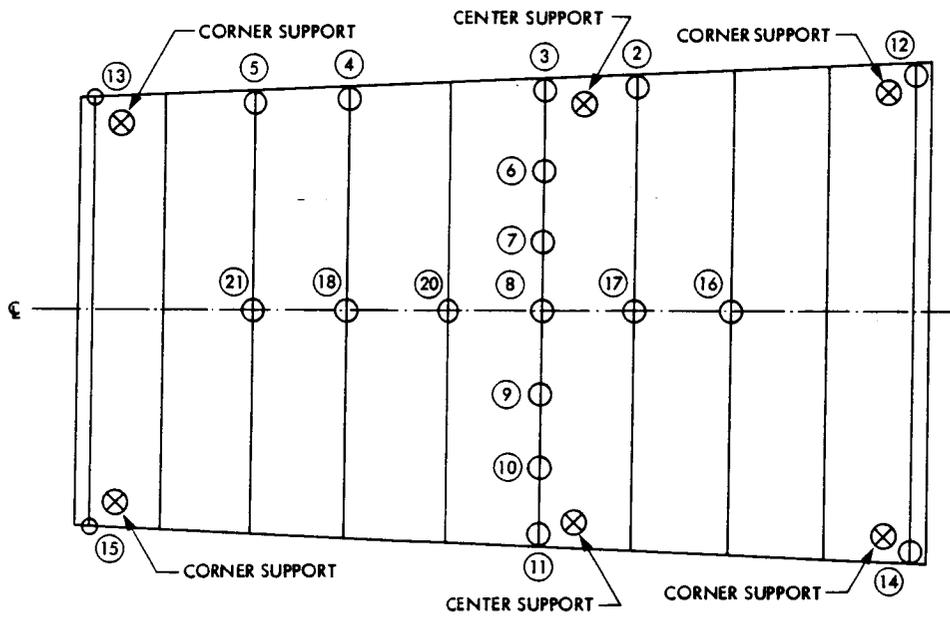


Fig. 6. Locations of the dial indicators



Fig. 7. Distributed load test using sandbags



Fig. 8. Up to 84 sandbags were used to represent the distributed load

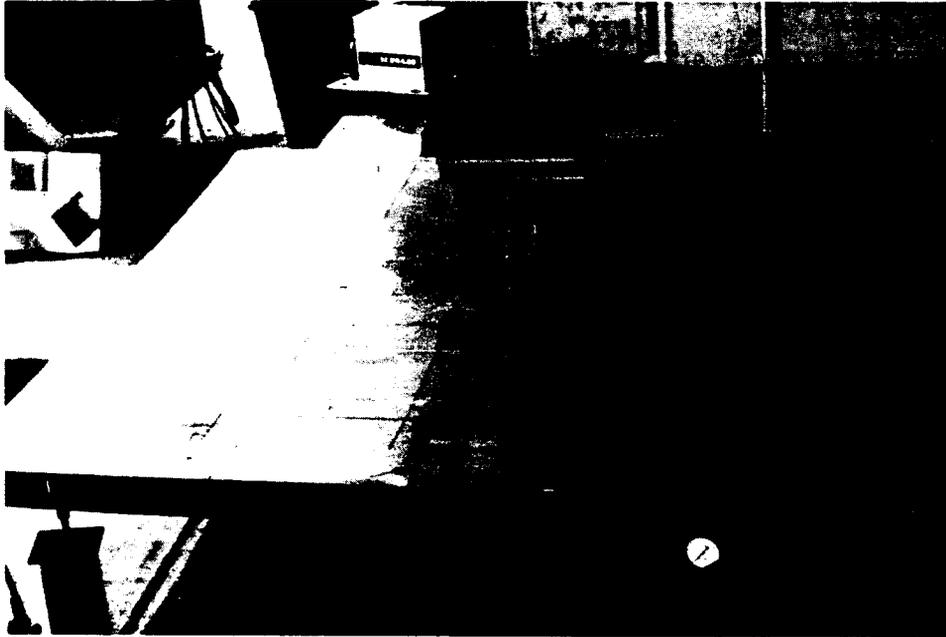


Fig. 9. Steel plates were placed on top of the panel to simulate the concentrated load

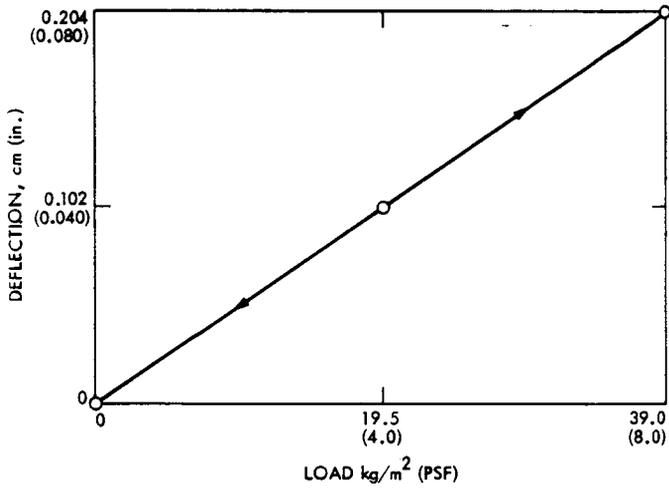


Fig. 10. Panel deflection at dial indicator number 8 as a function of the distributed load (panel without center supports)

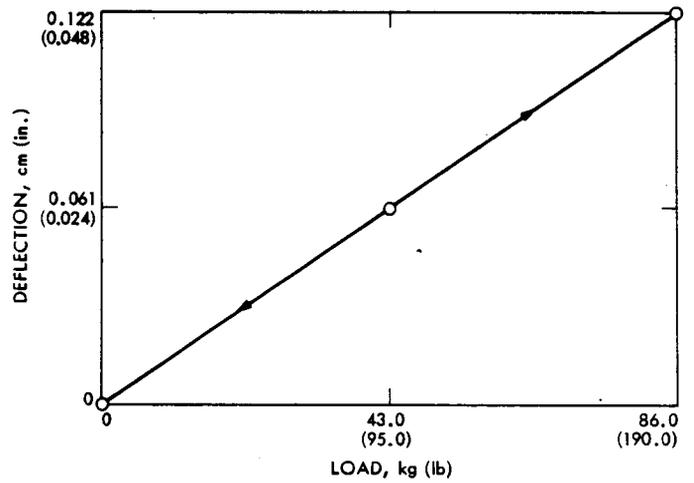


Fig. 11. Panel deflection at dial indicator number 17 as a function of the concentrated load (panel without center supports)

Appendix A

Computer Program for Panel Deflection Analysis (PANELDEFL)

I. Program Function

This computer program determines the deflections for one or a number of paraboloidal microwave antenna-reflector surface panels. In addition to performing the deflection analysis, each panel is best fit to minimize the mean square deflection errors from its ideal surface. Three fitting parameters are determined by a least squares analysis. These parameters consist of a shift in the coordinate normal to the plane that approximates the panel surface, and the two independent rotations about the panel's X- and Y-coordinate axes, which are as shown on Fig. A-1. Independently best-fitting individual panels is not always a valid procedure, so that the mean, the root-mean square (rms), and the standard deviation of the panel errors are supplied both with and without the fit. These are furnished both for the direction approximately normal to the panels' surface and also for the microwave pathlength direction. Weighting factors are applied to the analysis to approximate the area associated with each point of deflection calculation.

A comprehensive analysis of the entire set of antenna panels can be obtained by modelling one panel from each of the annular rings of panels that form the antenna surface. When this is done, the deflection analysis for each ring is synthesized to provide the rms microwave half-pathlength error for the entire antenna's panels. If the external loadings on the panels are specified to be the gravity load (weights) of each panel, then the synthesized deflections for each ring are assembled to provide the rms error for the entire surface at different antenna elevation angles of 0, 30, 60, and 90 degrees. These results are also recomputed with a common shift in the antenna focal axis direction of the entire set of panels to minimize the assembled surface error. When entire antenna surface calculations are made, each panel ring is weighted by its area relative to the entire aperture area.

II. Panel Configuration and Model

For deflection analysis each panel comprises a pair or radial girders that support a number (NRING)¹ of approximately circumferential (actually chordal) beams as shown in Fig. A-1. A number (NRIB) of equally spaced points along the axis of each beam can be specified at which deflection

¹Capitalized terms within parenthesis are FORTRAN namelist input parameters supplied by the user. See the definitions following the \$INPUT in the user instruction section.

calculations will be made. The first and the last of these points will actually fall on the girders. A trapezoidal configuration for each panel is determined by the inner (RI) and outer (RO) radii of the panel location within the antenna and by the central angle (TH). For RF pathlength analysis, it is necessary to specify the focal length (F) of the paraboloidal surface, or an approximating focal length of a quasi-paraboloidal surface.

Elastic properties of the structure are determined by specifying the bending moment of inertia of the girders (GIRDI) and beams (BEAMI), the modulus of elasticity (E) and bending efficiency factors for the girders (GFACT) and beams (BFACT). These factors are usually less than unity because of the effects of slitting the girders and beams in the fabrication process or because of other effects, such as web shear deflections. The factors are best determined by test. They conceivably could be greater than unity because of integral combining action of the surface skin sheet with the girder and beam cross sections. The surface skin is assumed to be parasitic and not to contribute to the strength of the panel.

The typical panel is assumed to be supported at two points on each girder at a distance (ENDG) in from the end of the panel. The axes of the girders are inward a distance (ENDB) from the radial edges of the panel. Thus, each beam span is the width of the trapezoid at the beam axis less twice the girder edge distance. The girders can also be specified to have an interior (ICENT) redundant support, which can be located by several methods (RINT).

Although not used in deflection calculations, the weight of each panel is computed from the cross-sectional areas of the beam (AB), girders (AG), panel skin thickness (THICK), and the density (DENS). A weight breakdown is given to show the contributions of each of these three types of components as well as the weight per unit of planform area.

III. Deflection Analysis

The loading for deflection analysis is a uniform surface load (PSF) to simulate either a wind pressure loading or a gravity loading. If the gravity loading is not known *a priori*, it is reasonable to run all the other data through the program just for the purpose of obtaining panel weight to establish the gravity loading.

The deflection analysis is made for beams and girders by integrating the usual differential equation that equates the

bending moment to the second derivative of displacement with respect to the longitudinal coordinate. The axes of beams and girders are assumed to be straight and contained in the secant plane. The secant plane is the plane that contains all four corner (or near-corner) girder supports. The deflections are computed normal to this plane. The loads on the beams are the distributed line loads determined from the surface loading and the beam spacing and are applied normal to the secant plane. The girder loads are concentrated reactions of the beams. The case of the redundant interior girder support is solved by the method of consistent displacements; the girder deflection is determined for a unit load at the redundant support and the unit load is scaled and applied to annihilate the girder deflection that is computed from the beam reactions when ignoring the interior support.

User Instructions

Runstream for JPL UNIVAC 1100/80 E computer:

@XQT 52219*RIL.PANELDEFL/MAP.

\$\$INPUT One set of namelist data per representative panel for a maximum of 24 rings.

\$\$END

.
.

.

.

.

\$\$INPUT Last panel.

\$\$END

@EOF

The NAMELIST input data are supplied for each panel ring as follows (see Fig. A-1 for sketch showing geometric variables):

F Focal length of parent parabola.
 RI Inner radius of the panel ring projected on the aperture plane.
 RO Outer radius of the panel ring projected as above.
 NRING Number of rings within each panel. This is also the number of circumferential beams, including the beams at RO and RI. Maximum = 30.
 NRIB Number of radial lines within the panel. This sets the number of points on each ring beam for deflection calculations. The first and last lines are on the panel radial edge girders.

TH Central angle of each panel in the ring, degrees.
 AB Cross-sectional area of the beams.
 AG Cross-sectional area of the girders.
 GIRDI The bending moment of inertia of the girders.
 BEAMI The bending moment of inertia of the beams.
 GFACT Effective stiffness factor for girders.
 Default = 0.75.
 BFACT Effective stiffness factor for beams.
 Default = 0.75.
 ICENT If not zero the girder has an interior support in addition to the end supports.
 RINT Distance to the interior girder support (for ICENT not zero).
 If RINT = 0.0: Support will be at the girder's center.
 If RINT = positive: It is the radial distance from the center of the antenna, measured perpendicular to the focal axis (same coordinate system as for RO, RI).
 If RINT = negative: It is the slope distance along the girder starting from RI.
 DENS Material density.
 E Young's Modulus.
 PSF Superimposed loading normal to the surface of the panel in pounds per square foot. This is the only loading for which deflection analysis is made and this could be different for each panel ring.
 THICK Average thickness of panel skins (allow for perforation) used for weight calculations.
 SLOPE The secant slope of the panel in the parent parabola. This is computed and echoed by the program. The loading is applied normal to this slope.
 IBUG 0 = minimum printout.
 3 = maximum printout.
 ENDG Distance from the end of the girder to the corner support point.
 ENDB Distance from the center line of the girder to the side edge of the panel.
 \$\$END

NOTES:

1. The minimum printout consists of the deflection matrix of the panel, panel weight, and summary rms analysis for each panel. For rms analysis, interior points are weighted by unity, edge points by one-half, and edge corner points by one-quarter.
2. Although the panel weight (beams plus girders plus skin) is computed, deflection analysis is made only for the uniform loads defined by PSF.
3. The load for deflection analysis is applied normal to the approximating plane of the panel skin. If PSF is made equal to the gravity weight of the panel then the associated deflections can be considered as the built-in bias at manufacturing. At specific antenna elevations (0°, 30°, 60° and 90°) the composite pathlength error is computed for all panel rings

and for the variable spring-back deflection for each of the panels in a full 360° panel ring. For the composite analysis of the entire surface each antenna ring is weighted by its area.

4. A best fit rms pathlength analysis is also determined for the composite antenna on the basis of a shift in the antenna Z (axial) coordinate.
5. The girder effective stiffness is computed as $GFACT * GIRDI * E$. Similarly the beam stiffness is $BFACT * BEAMI * E$.
6. Unless noted above, units are customary English inches and pounds.
7. The namelist data for successive panels does not need to repeat any information for prior panels that continue to be applicable.

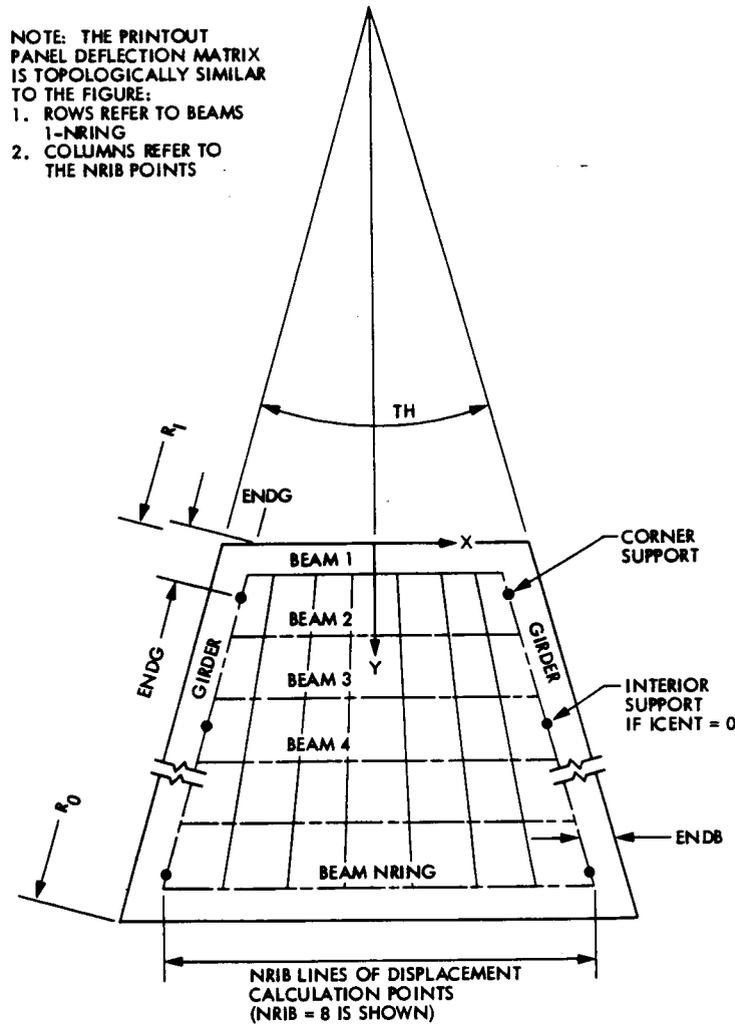


Fig. A-1. Panel configuration

Appendix B

Panel Structural Model by NASTRAN Program

A panel structural model was developed by using the NASTRAN (NASA Structural Analysis) finite element program for the tested panel.

The essential input data and procedure are summarized in Fig. B-1.

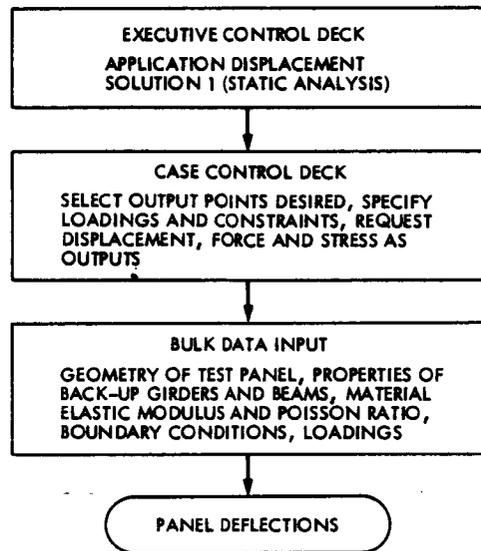


Fig. B-1. Flowchart of the essential input data and procedure

A Sequential Decoding Performance Analysis for International Comet Explorer

D. Divsalar

Communications Systems Research Section

This article analyzes the effect of a noisy reference carrier on the performance of International Comet Explorer (ICE) Sequential decoder. Ideal reference models for frame deletion probability are given for the Telemetry Processing Assembly (TPA) and the Linkabit (LS4815) sequential decoders. Based on these ideal reference models the deletion probability in the presence of noisy reference carrier is computed for both the high and the low data rate cases. A medium data rate performance model is then derived using an interpolation method. The derived medium rate performance model depends on the phase locked loop bandwidth-integration time product. Results are obtained for frame length of 1024 bits, data bit rates of 1024 bps and 512 bps, modulation index of 60° , and threshold loop bandwidths of 10 Hz and 3 Hz. The medium rate model agrees with experimental data to within measurement uncertainty. For the 10 Hz loop bandwidth and modulation index of 60° at frame deletion rate of 10^{-4} , the required total received power to noise ratio is 36 dB.

The analysis given in this article is general and applicable to any sequential decoder, provided that the ideal reference model of the decoder is available.

I. Introduction

Sequential decoding is a useful technique for communicating at low undetected error rates from deep space probes. But a failure mode known as computational overflow or erasure limits the operation of the decoder at very high data rates. The erasure (deletion) of a data frame occurs when the decoder has not finished decoding that frame at the time that it must be output. Increasing the speed of the decoding computation and the size of the decoder buffer provides some improvement in performance, but only linearly. On the other hand, if the buffer size is increased too much, erasures occur in bursts rather than singly. Such bursts decrease the expected improvement. Also if the speed of decoding is increased too

much with respect to data rate, then the decoder idle time waiting for incoming data will be increased. Therefore much of the increased capability is wasted.

Based on the received symbol sequence and past decoded bits, the sequential decoder makes local best estimates of the current data sequence. If the received symbols are relatively noiseless, decoding proceeds rapidly with no searching. On the other hand, if the received symbols are noisy, some of the local estimates of current data will be wrong. The decoder must eventually recognize that an error has been made and search systematically backward through the local estimates to correct those in error. The amount of searching depends on the amount of noise in the received symbol sequence.

In this article, the effect of receiver phase-locked loop (PLL) phase jitter on the frame deletion probability P_d of the sequential decoder is modeled for the International Comet Explorer (ICE) link. The approach is first to obtain an ideal-reference (baseline) model for the deletion probability of the sequential decoder. This has been obtained using models and simulation results in Ref. 1. These simulation and measurement results do not completely match our problem. Therefore approximation was used for the ideal-reference models. For the Linkabit (LS4815) sequential decoder, there are only two measurement points available. So the ideal-reference model used for LS4815 is also an approximation. Using these baseline models, first the expected deletion probability is calculated by averaging over the deletion probability conditioned on the PLL phase error, with the assumption that the PLL phase error is varying sufficiently slowly over a frame. The expected deletion probability, under this assumption, serves as an upper bound on the actual deletion probability and will be called the high data rate model for frame deletion probability. Then a lower bound on the deletion probability is obtained, which corresponds to the situation when the PLL phase error is varying rapidly over one-bit time. For this case the effective SNR has been computed and is inserted in the ideal reference model. The result is called the low data rate model. With these extreme models, an interpolation method similar to that in Refs. 2 and 3 has been used to interpolate a more realistic performance approximation between these bounds which is called the medium data rate model.

II. System Model

The ICE telemetry system is shown in Fig. 1. The convolutional code is rate 1/2, the constraint length $K = 24$, and the data frame length 1024 bits. The output symbol sequence of the encoder is Manchester coded before it phase modulates the carrier. At the receiver, the channel noise with two sided power spectral density $N_0/2$ is added to the received signal. Let P_T be the total received power. Then P_T/N_0 represents the total received power to noise spectral density ratio. Throughout the deletion probability calculation, we assume nominal 0.5 dB degradation loss due to signal conditioning, symbol synchronization and Manchester decoding assembly.

III. Ideal Reference Frame Deletion Probability Models

Based on simulation results in Ref. 1, the frame deletion probability has been approximated by a function for Helios frame size of 1152 bits (Refs. 1 and 4) as

$$P_d = \exp \left\{ \sum_{i=1}^3 \sum_{j=1}^3 A_{ij} R^{i-1} (\rho n N)^{j-2} \right\} \quad (1)$$

where matrix A with coefficients A_{ij} is given as

$$A = \begin{bmatrix} 2.397 & 8.824 & -0.9887 \\ -0.5331 & -6.788 & 1.569 \\ 0.02303 & 0.8848 & -0.8543 \end{bmatrix}$$

In Eq. (1), $R = E_b/N_0$ is the bit signal-to-noise ratio and N is the average number of computations per bit. If the computational speed of the decoder is C computations/s, the decoder buffer size is B bits, the frame size is F and the information rate is r_b bps, then

$$N = \frac{C \cdot B}{F \cdot r_b} = \frac{C \cdot B \cdot T_b}{F} \quad (2)$$

where T_b is the bit time.

In order to find noisy reference deletion probability, we should first find the ideal reference (perfect carrier reference) frame deletion probability $P_d(R)$. Using as a model Eq. (1) and noting that ICE frame length size is close to Helios frame length of 1152, with some approximation the following models for frame deletion for TPA and LS4815 decoders are proposed:

$$P_d(R) = \exp \{ \alpha_0 + \alpha_1 R + \alpha_2 R^2 \} \quad (3)$$

where α_0 , α_1 and α_2 are given in Table 1 for various cases. To find the noisy reference medium data rate model for deletion probability, we first find two extreme cases, namely the high data rate model and the low data rate model.

IV. The High Data Rate Model

For the high data rate model we assume that the carrier phase error is constant during one frame computation. Therefore the expected frame deletion probability can be defined by

$$P_{dh} = \int_{-\pi}^{\pi} P_d(R \cos^2 \phi | \phi) p(\phi) d\phi \quad (4)$$

where $P_d(R \cos^2 \phi | \phi)$ is the conditional deletion probability, for a given phase error ϕ , which can be found by replacing R with $R \cos^2 \phi$ in Eq. (3), the ideal reference deletion probability model. In Eq. (4), the $p(\phi)$ is the PLL phase error probability density, which is given (Ref. 5) approximately by

$$p(\phi) = \frac{\exp \{ \rho \cos \phi \}}{2\pi I_0(\rho)} \quad |\phi| \leq \pi \quad (5)$$

where ρ is the PLL signal-to-noise ratio, given by (Ref. 5)

$$\rho = \frac{P_T \cos^2 \theta}{N_0 B_L \Gamma_C} \quad (6)$$

and $I_0(\cdot)$ is the modified Bessel function of order zero. In Eq. (6) θ is the modulation index; B_L is the PLL bandwidth given by (Ref. 5)

$$B_L = B_{LO} \left(\frac{1 + r_0 \alpha / \alpha_0}{1 + r_0} \right) \quad (7)$$

where B_{LO} is the loop bandwidth at threshold, r_0 is the damping parameter at threshold, and α is the loop suppression factor given by (Ref. 5)

$$\alpha = \sqrt{\frac{0.7854 \rho_{in} + 0.4768 \rho_{in}^2}{1 + 1.024 \rho_{in} + 0.4768 \rho_{in}^2}} \quad (8)$$

and ρ_{in} is the input signal-to-noise ratio to the bandpass limiter and an IF filter having bandwidth B_{IF} . Defining the Carrier Margin (CM) by

$$CM = \frac{P_T \cos^2 \phi}{N_0 2 (B_{LO})} \quad (9)$$

then ρ_{in} can be expressed as

$$\rho_{in} = CM \times \rho_0 \quad (10)$$

with

$$\rho_0 \triangleq \frac{2B_{LO}}{B_{IF}} \quad (11)$$

In Eq. (7) α_0 is loop suppression factor evaluated by Eq. (8) with $\rho_{in} = \rho_0$. Finally Γ_C in Eq. (6) is the limiter performance factor given by (Ref. 5)

$$\Gamma_C = \frac{1 + \rho_{in}}{0.862 + \rho_{in}} \quad (12)$$

V. The Low Data Rate Model

For the low data rate model, we assume that the carrier phase error is varying rapidly during the bit time. This is true if loop bandwidth-bit time product is much larger than 1, i.e.,

$$B_L T_b \gg 1 \quad (13)$$

To compute degraded deletion probability for this case we should replace the ideal reference bit signal-to-noise ratio R in Eq. (3) by the effective bit signal-to-noise ratio $R \bar{x}^2$, where \bar{x} is given by

$$\bar{x} = E \{ \cos \phi \} = \frac{I_1(\rho)}{I_0(\rho)} \quad (14)$$

In Eq. (14), $I_1(\cdot)$ is the modified Bessel function of first order. Then we can get the low data rate model for frame deletion probability as

$$P_{d\ell} = P_d (R \bar{x}^2) \quad (15)$$

where $P_d(\cdot)$ is given by Eq. (3).

VI. The Medium Data Rate Model

In sequential decoding, when the data rates are roughly between $2B_{LO}$ and $2B_{LO}F$ bps, they should be considered as medium data rates. Therefore the performance of the sequential decoder under the noisy reference for the medium data rates should lie somewhere between the high data rate and the low data rate performance models. In order to have an accurate medium data rate model, we should predict the effective memory duration or the effective integration time for the decoder to make a decision on each decoded bit. Let T_m denote the average integration time (time required on average to test different branches through the tree diagram). It is true that

$$T_b \leq T_m \leq FT_b \quad (16)$$

But if a long search occurs, the search pattern is more likely to look like a full tree search. In Ref. 1 the effective integration time T_m has been found to be well approximated by

$$T_m = 2T_b \left(1 - \frac{1}{N} \left(\log_2 \left(1 + \frac{N}{2} \right) \right) \right) \quad (17)$$

The following computation has been used in Ref. 1 to derive Eq. (17). If ϕ_n denotes the number of tests required in a full binary tree with branch depth of n , then ϕ_n can be found from recursion

$$\phi_n = 1 + 2 \phi_{n-1} \quad (18)$$

The solution to Eq. (18) is

$$\phi_n = 2^n - 1 \quad (19)$$

Similarly if we consider both forward and backward moves contributing to a search of N steps then the branch depth n is the solution to the equation

$$\phi_n = \frac{N}{2} \quad (20)$$

or

$$n = \log_2 \left(1 + \frac{N}{2} \right) \quad (21)$$

On the other hand the accumulated length of tested branches θ_n can be found from recursion

$$\theta_n = n + 2\theta_{n-1} \quad (22)$$

The solution to Eq. (22) is

$$\theta_n = 2^{n+1} - n - 2 \quad (23)$$

Then

$$T_m = \frac{\theta_n}{\phi_n} T_b \quad (24)$$

where n satisfies Eq. (20) or (21). Using Eqs. (19), (21) and (23) in Eq. (24) results in Eq. (17). For large N from Eq. (17)

$$T_m \approx 2 T_b \quad (25)$$

As follows, we represent the medium data rate deletion probability model as an interpolation between the high and the low data rate models. We will find the interpolation parameter in terms of loop bandwidth-integration time product $B_L T_m$. This approach is similar to a technique used for uncoded telemetry and block codes (Ref. 2). Let's define the conditional effective bit signal-to-noise ratio as

$$R_{\text{eff}} = R x^2 \quad (26)$$

where

$$x \triangleq \frac{1}{T_m} \int_0^{T_m} \cos \phi(t) dt \quad (27)$$

Note that for $B_L T_m \gg 1$ the phase error $\phi(t)$ changes rapidly over a bit time; thus

$$x \xrightarrow{B_L T_m \rightarrow \infty} \bar{x} = E \{ \cos \phi \} \quad (28)$$

where we have assumed $\phi(t)$ is a stationary ergodic process. When $B_L T_m \ll 1$, $\phi(t)$ remains essentially constant over the integration time T_m and thus

$$x \xrightarrow{B_L T_m \rightarrow 0} \cos \phi \quad (29)$$

Clearly these two extreme cases result in the low data and the high data rate models for deletion probabilities, respectively. Now for anything between these two extremes we can write

$$P_{dm} = \int P_d(Rx^2 | x) p(x) dx \quad (30)$$

where x is defined by Eq. (27) and $p(x)$ is the probability density function of the random variable x . The approximate probability density function of x is given in Ref. 1, which enables us to evaluate P_{dm} from Eq. (30). But rather than doing so, we prefer to find the high and low data rate models by using more accurate phase error distributions and then finding the interpolation parameter approximately. Using the ideal reference frame deletion probability models in Eq. (3), we have

$$P_d(Rx^2 | x) = \exp \{ \alpha_0 + \alpha_1 Rx^2 + \alpha_2 R^2 x^4 \} \quad (31)$$

Using the Taylor's series expansion of Eq. (31) around $\bar{x} = E \{ \cos \phi \}$, we obtain

$$P_d(Rx^2 | x) = P_d(R\bar{x}^2) (1 + \beta_1 (x - \bar{x}) + \beta_2 (x - \bar{x})^2 + \dots) \quad (32)$$

where

$$\beta_1 = 2 \alpha_1 R \bar{x} + 4 \alpha_2 R^2 \bar{x}^3 \quad (33)$$

and

$$\beta_2 = \alpha_1 R + \alpha_1 R \bar{x} + 6 \alpha_2 R^2 \bar{x}^2 + 2 \alpha_2 R^2 \bar{x}^3 \quad (34)$$

Noting that $P_{d\ell} = P_d(R\bar{x}^2)$ and using Eq. (32) in Eq. (30) we get

$$P_{dm} = P_{d\ell} (1 + \beta_2 \sigma_x^2 + \dots) \quad (35)$$

Also if we use Eq. (32) in Eq. (30) with assumption that $B_L T_m \ll 1$, then we get

$$P_{dh} = P_{d\ell} (1 + \beta_2 \sigma_{\cos\phi}^2 + \dots) \quad (36)$$

From Eqs. (35) and (36), we conclude that if σ_x^2 and $\sigma_{\cos\phi}^2$ are small and if other central moment terms are ignored, we obtain the required interpolation formula as

$$P_{dm} = \left(1 - \frac{\sigma_x^2}{\sigma_{\cos\phi}^2}\right) P_{d\ell} + \frac{\sigma_x^2}{\sigma_{\cos\phi}^2} P_{dh} \quad (37)$$

At this point we model $\phi(t)$ as a Gaussian process having the same variance and bandwidth as the actual $\phi(t)$ process and with an autocorrelation function derived from a first-order PLL (Ref. 4):

$$R_\phi(\tau) = \sigma_\phi^2 \exp\{-4B_L |\tau|\} \quad (38)$$

Hence,

$$\begin{aligned} \sigma_x^2 &= E(x - \bar{x})^2 \\ &= \frac{1}{T_m^2} \int_0^{T_m} \int_0^{T_m} R_{\cos\phi - \bar{x}}(t_1 - t_2) dt_1 dt_2 \end{aligned} \quad (39)$$

where

$$R_{\cos\phi - \bar{x}}(t_1 - t_2) = E\{(\cos\phi(t_1) - \bar{x})(\cos\phi(t_2) - \bar{x})\} \quad (40)$$

Using the approximation

$$\cos\phi(t) \approx 1 - \phi^2(t)/2 \quad (41)$$

we get

$$R_{\cos\phi - \bar{x}}(t_1 - t_2) \approx \frac{[E\{\phi^2(t_1)\phi^2(t_2)\} - \sigma_\phi^4]}{4} \quad (42)$$

Since we modeled $\phi(t)$ as a Gaussian process with the auto-correlation given in Eq. (38), we can easily find the joint density function of $\phi(t_1)$ and $\phi(t_2)$. Having this joint density function, we can derive the following:

$$E\{\phi^2(t_1)\phi^2(t_2)\} = \sigma_\phi^4 + 2R_\phi^2(t_1 - t_2) \quad (43)$$

and

$$R_{\cos\phi - \bar{x}}(t_1 - t_2) \approx \frac{1}{2} R_\phi^2(t_1 - t_2) \quad (44)$$

Note that

$$\sigma_{\cos\phi}^2 = R_{\cos\phi - \bar{x}}(0) \approx \frac{1}{2} R_\phi^2(0) \quad (45)$$

Using Eq. (44) in Eq. (39) with Eq. (45) we get

$$\begin{aligned} \frac{\sigma_x^2}{\sigma_{\cos\phi}^2} &\approx \frac{1}{T_m^2} \int_0^{T_m} \int_0^{T_m} \exp\{-8B_L |t_1 - t_2|\} dt_1 dt_2 \\ &= \frac{1}{4B_L T_m} - \frac{1}{32B_L^2 T_m^2} (1 - \exp(-8B_L T_m)) \end{aligned} \quad (46)$$

Substituting Eq. (46) in Eq. (37), we have the medium data rate frame deletion probability P_{dm} .

VII. Conclusion and Numerical Results for ICE

In this article, we have obtained models for ideal reference frame deletion probabilities, based on results in Ref. 1, for TPA and LS4815 sequential decoders. Then we have derived the high and low data rate models for noisy reference frame deletion probabilities. Finally, using an interpolation method, we derived the medium data rate model for noisy reference frame deletion probability. The results are shown in Figs. 3 through 8. In Fig. 2, the PLL bandwidth vs P_T/N_0 is depicted. In Figs. 3 through 6 P_d vs P_T/N_0 for various models is shown for two-sided threshold loop bandwidth of 10 Hz and 3 Hz, and for bit rates 1024 and 512 bps, for TPA decoder. In Figs. 7 and 8, the results are given for the LS4815 sequential decoder with 1024 bps. These theoretical models are in good agreement with measurement results (J.W. Layland, private communication).

Acknowledgment

The author wishes to thank J. H. Yuen of Sec. 331 and J. W. Layland of the Telecommunications and Data Acquisition Office for their useful comments and suggestions.

References

1. Layland, J. W., "A sequential decoding medium rate performance model," in *The Deep Space Network Progress Report*, Jet Propulsion Laboratory, Pasadena, CA, Dec. 15, 1973, Tech. Rep. 32-1526, vol. XVIII, pp. 29-40.
2. Tausworthe, R. C., "Communications Systems Development: Efficiency of Noisy Reference Detection," in *The Deep Space Network Progress Report*, Jet Propulsion Laboratory, Pasadena, CA, Dec. 15, 1968, Space Programs Summary 37-54, vol. III, pp. 195-201.
3. Lesh, J. R., "Sequential Decoding in the Presence of a Noisy Carrier Reference," *IEEE Trans. Communications*, Vol. COM-23, Nov. 1975, pp. 1292-1297.
4. Yuen, J. H., *Deep Space Telecommunications Systems Engineering*, Plenum Press, New York, Oct. 1983.
5. Lindsey, W. C. and Simon, M. K., *Telecommunications Systems Engineering*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.

Table 1. Model coefficients for frame-deletion rate

	C , Computational speed, comp/sec	B , Buffer size, bits	r_b , Bit rate, bps	α_0	α_1	α_2
TPA	92,500	3×1024	1024	3.71	1.9	-3.89
	92,500	3×1024	512	2.98	3	-4.49
LS4815	10^6	4×1024	1024	0.936	6.12	-6.177
	10^6	4×1024	512	0.23	7.2	-6.77

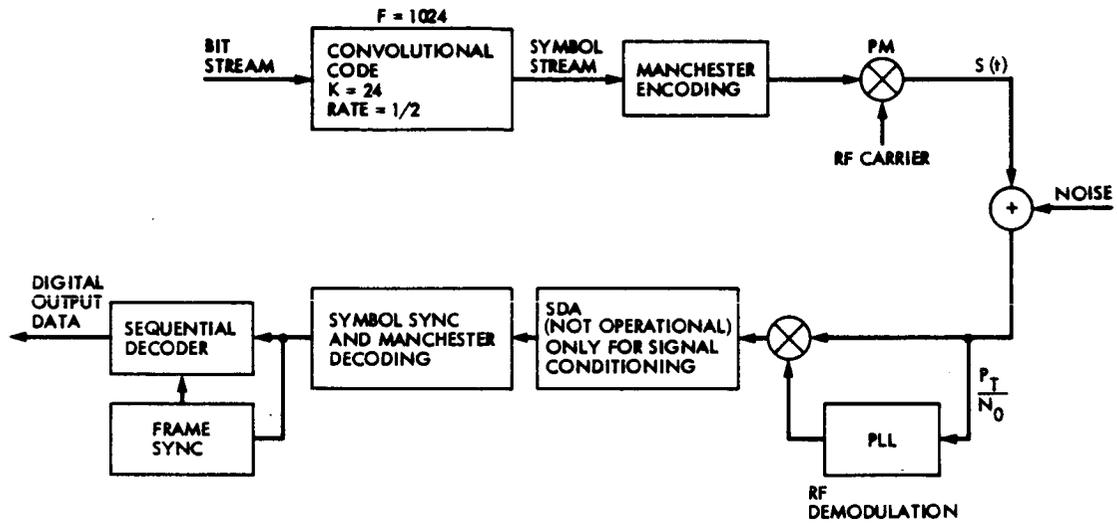


Fig. 1. Telemetry link configuration

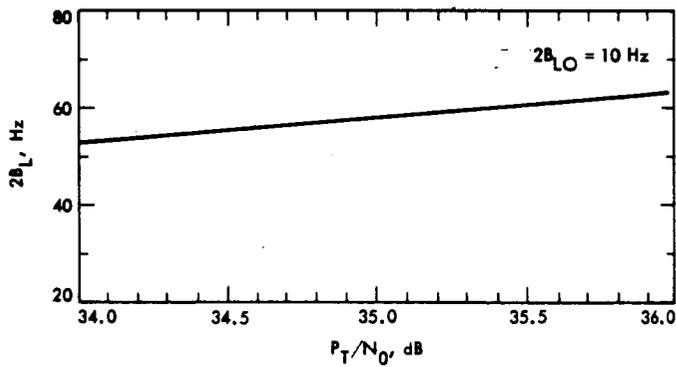


Fig. 2. PLL bandwidth vs P_T/N_0 for $2B_{LO} = 10$ Hz

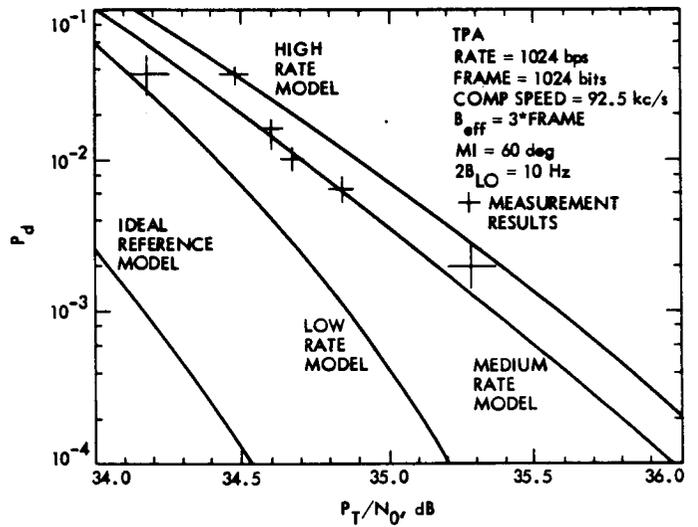


Fig. 3. Frame deletion probability models vs P_T/N_0 for TPA, bit rate = 1024 bps and $2B_{LO} = 10$ Hz

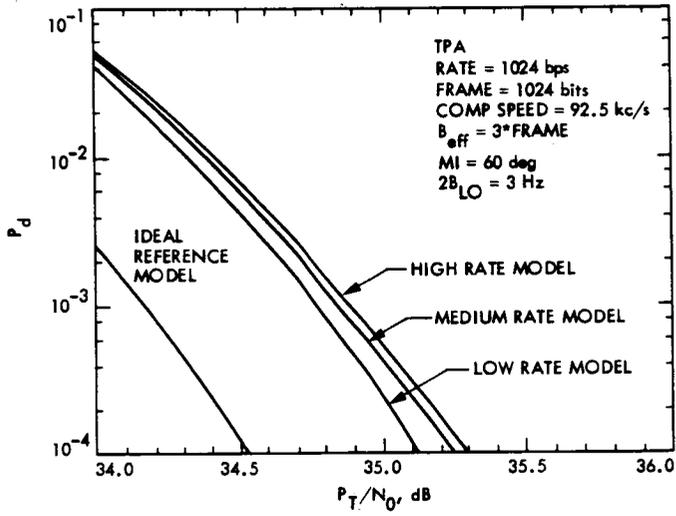


Fig. 4. Frame deletion probability models vs P_T/N_0 for TPA, bit rate = 1024 bps and $2B_{LO} = 3 \text{ Hz}$

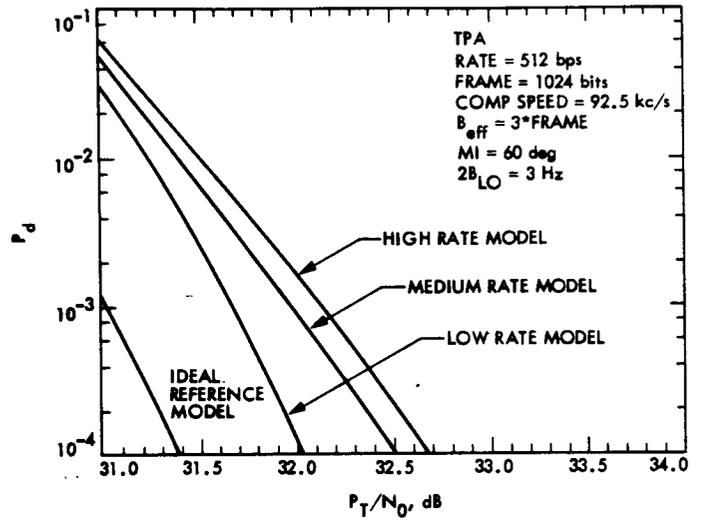


Fig. 6. Frame deletion probability models vs P_T/N_0 for TPA, bit rate = 512 bps and $2B_{LO} = 3 \text{ Hz}$

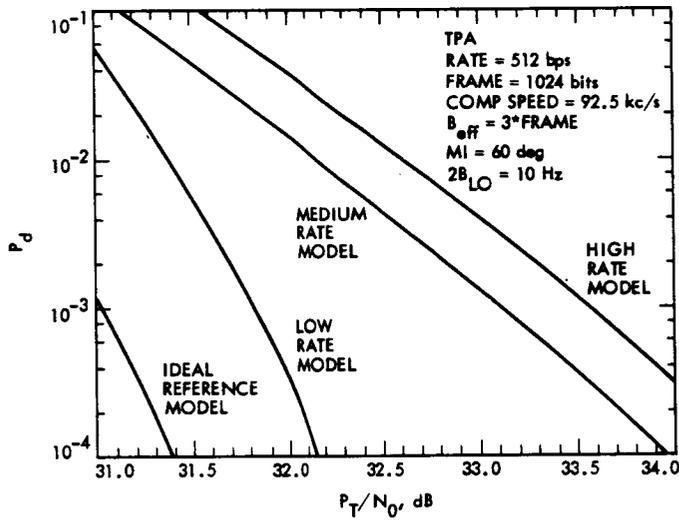


Fig. 5. Frame deletion probability models vs P_T/N_0 for TPA, bit rate = 512 bps and $2B_{LO} = 10 \text{ Hz}$

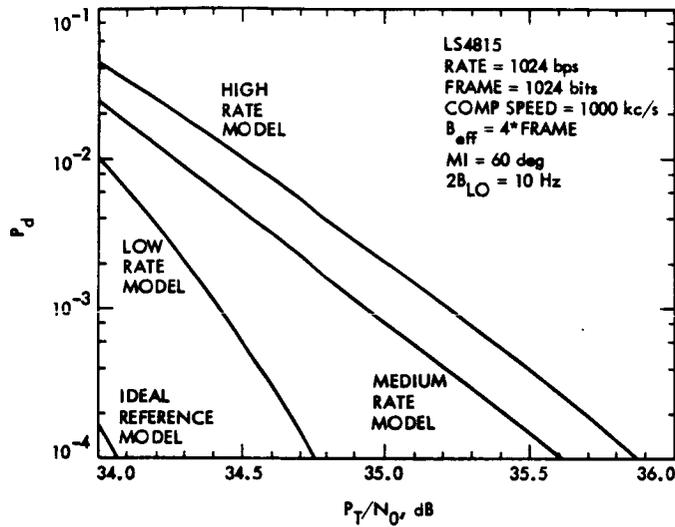


Fig. 7. Frame deletion probability models vs P_T/N_0 for LS 4815, bit rate = 1024 bps and $2 B_{LO} = 10 \text{ Hz}$

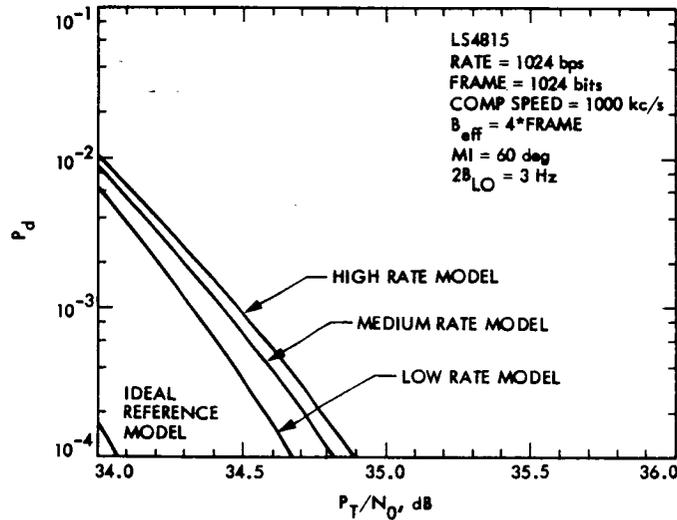


Fig. 8. Frame deletion probability models vs P_T/N_0 for LS 4815, bit rate = 1024 bps and $2 B_{LO} = 3 \text{ Hz}$

The VLSI Design of Error-Trellis Syndrome Decoding for Convolutional Codes

I. S. Reed and J. M. Jensen

University of Southern California, Los Angeles

T. K. Truong and I.S. Hsu

Communications System Research Section

In this article, a recursive algorithm using the error-trellis decoding technique is developed to decode convolutional codes (CCs). An example, illustrating the VLSI architecture of such a decoder, is given for a dual-K CC. It is demonstrated that such a decoder can be realized readily on a single chip with NMOS technology.

I. Introduction

Recently, the authors (Refs. 1, 2) developed a new error-trellis syndrome decoding scheme for convolutional codes (CCs). This new method involves finding minimum-error paths in an error-trellis. It was shown (Ref. 1) that the computation of the error-trellis is accomplished by finding the solution of the syndrome equations explicitly in terms of the actual error sequence. This syndrome decoding scheme was then applied to a rate 3/4, one-error-correcting systematic Wyner-Ash code (Refs. 4, 5).

In this article, the error-trellis decoding is applied to decode a rate 1/n, dual-K nonsystematic CC. The special example of a rate 1/2, dual-3 nonsystematic CC is treated in this article.

It is demonstrated in Ref. 6 and this article that the real advantage of error-trellis decoding over Viterbi decoding of CCs is the reduction of the number of states and transitions between any two frames. A recursive algorithm for finding the path of minimum error in the error-trellis is found which realizes a rate 1/2, dual-3 nonsystematic CC. This recursive algorithm eliminates all paths in certain fixed frames except

the path with minimum error. A VLSI chip architecture is developed to realize this new recursive algorithm for decoding the dual-K CC. The designs developed for this decoder are regular, simple, expandable and, therefore, naturally suitable for VLSI implementation.

II. The Properties of Convolutional Codes

In this section a brief review is presented of properties of CCs needed in the following sections.

The input \underline{X} to a k -dimensional CC encoder is represented as an infinite sequence of vectors, X_j , of form,

$$\underline{X} = [X_0, X_1, X_2, \dots] \quad (1)$$

where $X_j = [x_{1j}, x_{2j}, \dots, x_{kj}]$ is a k -component vector of elements from the Galois field $GF(q)$, where $q = p^n$ with p a prime integer. Each vector X_j of k symbols for ($j = 0, 1, 2, \dots$) is sometimes called an information or input frame (see Ref. 4, Sec. 12.1). Similarly, the output \underline{Y} of a CC of length n is an infinite sequence of vectors, Y_j , of form

$$\underline{Y} = [Y_0, Y_1, Y_2, \dots] \quad (2)$$

where $Y_j = [y_{1j}, y_{2j}, \dots, y_{nj}]$ is an n -component vector of elements from $GF(q)$. Here vector Y_j is called a codeword frame or more simply, a code frame (Ref. 4, Sec. 12.1).

In a CC encoder the input \underline{X} in Eq. (1) and output in Eq. (2) are linearly related; hence there exists what is called an infinite generator matrix \underline{G} such that

$$\underline{Y} = \underline{X} \cdot \underline{G} \quad (3)$$

For the CC to have finite memory, G has the form

$$\underline{G} = \begin{bmatrix} G_0 & G_1 & G_2 & \dots & G_m \\ & G_0 & G_1 & G_2 & \dots & G_m \\ & & & \vdots & & \\ & & & & & \vdots \\ & & & & G_0 & G_1 & G_2 & \dots & G_m \end{bmatrix} \quad (4a)$$

where the submatrices G_j are $k \times n$ matrices of form

$$G_j = [G_{rsj}] \quad (4b)$$

and the elements G_{rsj} belong to $GF(q)$ for $1 \leq r \leq k$, $1 \leq s \leq n$ and $0 \leq j \leq m$. Multiplying the subvectors of \underline{x} in Eq. (1) by the matrix G in Eq. (4a) yields, by Eq. (3), the fundamental identity,

$$Y_j = \sum_{i=0}^{\min(j,m)} X_{j-i} \cdot G_i \quad (5)$$

which is the convolution of sequence $\{X_0, X_1, \dots\}$ of information frames with the sequence $\{G_0, G_1, \dots, G_m\}$ of matrix operators. The integer m in Eq. (4a) is the memory of the convolution Eq. (5). The value of m is the maximum number of past input frames X_j needed to compute Eq. (5), recursively.

The convolution property (Eq. [5]) of finitely generated CCs can be realized conveniently for some applications by the operational calculus over a finite field $GF(q)$. To accomplish this one defines first the generating functions or, what are sometimes called, the D -transform of the sequences $\{X_j\}$, and the $\{G_j\}$ and $\{Y_j\}$ matrices, as follows:

$$X(D) = \sum_{j=0}^{\infty} X_j D^j \quad (6a)$$

$$g(D) = \sum_{j=0}^m G_j D^j \quad (6b)$$

and

$$Y(D) = \sum_{j=0}^{\infty} Y_j D^j \quad (6c)$$

where D is an operator variable. It is not difficult to verify that by equating the coefficients in the matrix relationship

$$Y(D) = X(D) \cdot G(D) \quad (7)$$

the fundamental convolution property (Eq. [5]) of a convolutional code of memory m is derived. Hence identity (Eq. [7]) is precisely equivalent to the defining relationship (Eq. [3]) of a convolutional code. Finally if D is identified with a unit delay circuit element, it is not difficult to show that $G(D)$ maps directly onto an encoder circuit diagram.

By Eq. (5) the j th output of an n -vector or codeword frame, Y_j is dependent on at most the $m+1$ present and past input k -vectors or information frames. Hence it is natural (as suggested by Blahut [Ref. 4, Section 12.1]) to define

$$k_1 = (m+1)k \quad (8a)$$

to be the word length of the CC. Then the word length k , is extended by the encoding process of Eq. (5) to, what is called, the block length n_1 of the CC. The block length of CC is

$$n_1 = (m+1) \cdot n = \frac{k_1}{R} \quad (8b)$$

where $R = k/n$ is the rate of the code. By Eq. (5) the block length $n_1 = (m+1)n$ is the length of the subsequence of \underline{Y} which, during encoding, can be influenced by a single information frame.

The minimum distance of the code is interpreted to be the Hamming weight of the smallest weight code word segment of $\ell = m+1$ which is nonzero in the first frame. Suppose for some CC that at most t errors occur during transmission in the first ℓ code word frames, and that

$$2t+1 \leq d$$

is satisfied by the code. Then those errors which occur in the first block length of CC can be corrected using feedback decoding. Such a CC is called a t -error-per-block-length-correcting CC or more simply a t -error-correcting CC (Ref. 4, Sec. 12.3).

Another distance between code words of a CC which is commonly used is the free distance d_{free} :

$$d_{\text{free}} = \min_{X(D) \neq 0} W_H(X(D) \cdot G(D))$$

Since clearly, $d \leq d_{\text{free}}$ (Ref. 4), designing a CC with minimum distance d guarantees that the code has a free distance of d or greater.

To find the minimum distance d of a CC, either the following $k_1 \times n_1$ submatrix is used

$$G = \begin{bmatrix} G_0 & G_1 & G_2 & \dots & G_m \\ & G_0 & G_1 & \dots & G_{m-1} \\ & & & \ddots & \\ & & & & G_0 & G_1 \\ & & & & & & G_0 \end{bmatrix} \quad (9)$$

where $k_1 = (m+1)k$ and $n_1 = (m+1)n$, or its corresponding parity check matrix H . The techniques used to find the minimum distance for block codes apply also for finding d using matrix G in Eq. (9) or the associated parity-check matrix H . Sometimes (see Ref. 4, Sec. 3.3) matrix G in Eq. (9) is called the basic generator matrix of the CC.

III. Error-Trellis Decoding

In this section error-trellis decoding as developed in Refs. 3 and 6 is reviewed briefly. First in order to avoid catastrophic error propagation the D -transform $G(D)$ in Eq. (6b) is restricted to have the Smith normal form

$$G(D) = A(D) [I_k, 0] B(D) \quad (10)$$

where $A = A(D)$ and $B = B(D)$ are, respectively, $k \times k$ and $n \times n$ invertible matrices with elements in $F[D]$, the ring of polynomials in D over $GF(g)$. The elements of the inverses A^{-1} and B^{-1} of matrices A and B , respectively, are also in $F[D]$ or are polynomials in D . For descriptive brevity the D -transform $G(D)$ is called the generator matrix.

If $B = B(D)$ in Eq. (10), let

$$B = [B_1, B_2]^T \text{ and } B^{-1} = [\overline{B}_1, \overline{B}_2] \quad (11)$$

where the first k rows of B constitute submatrix B_1 and the remaining $n-k$ rows are B_2 . Similarly the first k columns of B^{-1} constitute submatrix \overline{B}_1 and the other $n-k$ columns are \overline{B}_2 . It was shown (Refs. 1, 2, 6) that

$$G \cdot H^T(D) = G \cdot \overline{B}_2 = 0 \quad (12)$$

where $H(D)$ is a parity-check matrix for $G(D)$.

Let $Z(D) = [Z_1(D), \dots, Z_n(D)]$ be the vector D -transform of received sequence \underline{Z} . Then the D -transforms of the transmitted and received sequences are related by

$$Z(D) = Y(D) + e(D) \quad (13)$$

where $e(D) = [e_1(D), \dots, e_n(D)]$ is D -transform of error sequence, henceforth called, simply, the error sequence.

From Eq. (12) the syndrome of the received code word is

$$S = Z \cdot H^T = e \cdot \overline{B}_2 \quad (14)$$

This is a nonhomogeneous system of linear equations for the unknown error sequence $e(D)$. The problem of syndrome decoding CCs is to solve this system of equations for $e(D)$. It was shown (Refs. 1, 6) that the general solution of Eq. (14) is given by

$$e = uG + ZR \quad (15a)$$

where

$$R = \overline{B}_2 \cdot B_2 \quad (15b)$$

with B_2 and \overline{B}_2 defined in Eq. (11).

Let e_a replace e in Eq. (13) as the actual error sequence; then a substitution of Eq. (13) into Eq. (15a) gives by Eq. (12)

$$e = uG + ZR = uG + (Y + e_a) \overline{B}_2 B_2 = uG + e_a R$$

which is independent of the transmitted codeword Y .

Using the solution (Eq. (15a)) of the syndrome equation an estimate of the error sequence can be obtained from

$$\|\hat{e}\| = \min_u \|uG + ZR\| \quad (16)$$

where $\|\cdot\|$ denotes Hamming distance norm or weight and the minimization is taken over all k -vectors u over $F[D]$. By Eq. (16) the minimum weight error sequence is

$$\hat{e} = \hat{u}G + ZR = \hat{u}G + e_a R \quad (17)$$

where \hat{u} is the k -vector with elements in $F[D]$ for which the minimum weight in Eq. (16) is obtained. It was shown (Ref. 1)

that \hat{u} in Eq. (17) is a correction factor such that the original message is estimated by

$$\hat{X} = Z \cdot G^{-1} - \hat{u} \quad (18)$$

Substituting $Z = Y + e_a$ into Eq. (17) yields

$$\hat{X} = (Y + e_a) G^{-1} - \hat{u} = X + e_a G^{-1} - \hat{u} \quad (19)$$

Let E be the set of all error sequences which can be decoded correctly. Then, if $e_a \in E$, the most likely error sequence found by the minimization in Eq. (16) is equal to e_a , and therefore, by Eq. (19), $u = e_a G^{-1}$. Thus, the minimization in Eq. (16) has only to be taken over those sequences u which belongs to $E^{(-1)} = \{\hat{u} = e G^{-1} : e \in E\}$. Hence

$$\|\hat{e}\| = \min_{u \in E^{(-1)}} \|uG + ZR\| \quad (20)$$

Note that if $e_a \in E$, the most likely error sequence found by either Eq. (16) or (20) is identical and equal to e_a .

In order to actually perform the minimization in Eq. (20) over $E^{(-1)}$, the sets E and $E^{(-1)}$ must be identical. This is generally impossible. However for systematic CCs, it was shown (Ref. 6) that $E^{(-1)}$ can be approximated by the set

$$E_1^{(-1)} = \{u : W_H(u_j, \dots, u_{j+m}) \leq t, \text{ for all } j \geq 0\} \quad (21)$$

where $t = [(d_{\text{free}} - 1)/2]$ and m is the length of memory. For a more detailed discussion of the relation between $E^{(-1)}$ and $E_1^{(-1)}$, see Ref. 6. Thus, for systematic CCs, Eq. (20) becomes

$$\|\hat{e}\| = \min_{u \in E_1^{(-1)}} \|uG + ZR\| \quad (22)$$

In order to take the minimization in Eq. (22) over $E_1^{(-1)}$, a specific procedure was found to "prune" the error-trellis (Ref. 6). Also it was shown (Ref. 6) that the number of states S and transitions T needed for error-trellis decoding of an arbitrary systematic CC is

$$S = \sum_{i=0}^{\min\{t, mk\}} \binom{mk}{i} (q-1)^i \quad (23)$$

and

$$T = \sum_{i=0}^{\min\{t, (m+1)k\}} \binom{(m+1)k}{i} (q-1)^i \quad (24)$$

respectively. Note that the standard Viterbi decoding (Ref. 3, Sec. 4.17) requires q^{mk} states and $q^{(m+1)k}$ transitions within a frame time.

In the next section error-trellis decoding is developed for an important class of nonsystematic CCs, called dual- K CCs. The dual- K convolutional codes were invented and developed by Viterbi and Odenwalder. These CCs are nonbinary codes over the field $GF(2^K)$ and are used in practice in channels which experience fading such as UHF tropospheric scatter channels, etc.

IV. Error-Trellis Decoding of Dual- K CCs

Dual- K $(n, 1)$ convolutional codes are of rate $1/n$, of memory $m = 1$, and with symbols in the finite or Galois field $GF(2^K)$ (see Ref. 7). The generating matrix G is a special case of Eq. (4a), namely,

$$G^{(1)} = \begin{bmatrix} G_0 & G_1 & & & \\ & G_0 & G_1 & & \\ & & G_0 & G_1 & \\ & & & & \ddots \\ & & & & & \ddots \end{bmatrix} \quad (25a)$$

where

$$G_0 = [1, 1, 1, \dots, 1] \quad (25b)$$

$$G_1 = [g_{11}, g_{12}, \dots, g_{1n}]$$

with $g_{1j} \neq 0$ and $g_{1j} \in GF(2^K)$ and the g_{1j} 's are all distinct, for $1 \leq j \leq n$.

From the above definition of a dual- K CC, it is readily verified that the minimum distance of the code is $d = (2n - 1)$ and the free distance is

$$d_{\text{free}} = 2n \quad (25c)$$

Hence if no more than t symbol errors occur in the first 2 code word frames and $2t + 1 \leq d = 2n - 1$ or $t \leq n - 1$, then those errors which occur in the first frame can be corrected. In other words, the dual- K CC is a t -error-per-block-length-correcting CC where

$$t = \left\lfloor \frac{(d_{\text{free}} - 1)}{2} \right\rfloor = n - 1$$

and $\lfloor x \rfloor$ denotes the greatest integer less than x .

If error-trellis decoding is applied to the dual- K CCs, then from Eq. (20) the most likely error sequence \hat{e} is found as

$$\|\hat{e}\| = \min_{u \in E^{(-1)}} \|uG + ZR\| \quad (25d)$$

In Appendix A it is shown that $E^{(-1)}$ can be approximated by the set

$$\tilde{E}^{(-1)} = \{u: W_H(u_j, u_{j+1}) \leq t, \text{ for all } j \geq 0\} \quad (25e)$$

Thus error-trellis decoding of dual- K CCs is performed by taking the minimum in Eq. (25d) over the set $\tilde{E}^{(-1)}$ in Eq. (25e). But by Eq. (21) this set is equal to $E_1^{(-1)}$ which is used in Eq. (6) also for systematic CCs. Therefore, the trellis can be "pruned" using the procedure in Eq. (6), and also the number of states and transitions within a frame time is as given in Eqs. (23) and (24). Consider the example by Odenwalder (Ref. 7, Fig. 1).

Example 1. Let the Galois field $GF(2^3)$ be generated by the polynomial $x^3 + x^2 + 1$, irreducible over $GF(2^3)$. If α is a root of this polynomial, then

$$\begin{aligned} \alpha, \alpha^2, \alpha^3, \alpha^4 &= 1 + \alpha + \alpha^2, \alpha^5 = 1 + \alpha, \alpha^6 \\ &= 1 + \alpha, \alpha^7 = 1, \text{ and } 0 \end{aligned}$$

are the eight elements of $GF(2^3)$. The generating matrix of type Eq. (6b) for a rate $1/2$, dual-3 CC is

$$G = [1 + D, 1 + \alpha D] \quad (26a)$$

The output of encoder in terms of input is

$$Y = [Y_1, Y_2] = X[1 + D, 1 + \alpha D]. \quad (26b)$$

If one applies elementary column operations to G in Eq. (25b), it is not difficult to show that

$$G = [1, 0] \begin{bmatrix} 1 + D, & 1 + \alpha D \\ 1, & \alpha \end{bmatrix}$$

is the Smith normal from Eq. (10). Hence

$$B = \begin{bmatrix} 1 + D, & 1 + \alpha D \\ 1, & \alpha \end{bmatrix} \quad (27a)$$

$$B^{-1} = \begin{bmatrix} \alpha^3, \alpha^2 + \alpha^3 D \\ \alpha^2, \alpha^2 + \alpha^2 D \end{bmatrix} \quad (27b)$$

are the matrices needed in Eqs. (11a) and (15b).

$$\begin{aligned} R &= \overline{B_2} B_2 = \begin{bmatrix} \alpha^2 + \alpha^3 D \\ \alpha^2 + \alpha^2 D \end{bmatrix} [1, \alpha] \\ &= \begin{bmatrix} \alpha^2 + \alpha^3 D, \alpha^3 + \alpha^4 D \\ \alpha^2 + \alpha^2 D, \alpha^3 + \alpha^3 D \end{bmatrix} \end{aligned} \quad (28)$$

is the matrix R needed in the error-trellis solution, Eq. (15b), of the syndrome Eq. (14). From Eq. (25c), $d_{\text{free}} = 4$ and hence, the present dual-3 code will correct at most 1 symbol per block length of 2. From Eq. (24), the number of transitions in one frame time needed in the error-trellis is

$$T = \sum_{i=0}^1 \binom{(m+1)k}{i} (q-1)^i = \sum_{i=0}^1 \binom{2}{i} \cdot (8-1)^i = 15$$

For the standard decoding trellis $q^{(m+1)k} = 8^2 = 64$ transitions are required. This yields a fractional reduction of $15/64 \cong 1/4$ in the number of transitions needed for error-trellis decoding between that required for standard Viterbi hard decoding. Also from Eq. (23), the number of states is

$$S = \sum_{i=0}^1 \binom{mk}{i} (q-1)^i = \sum_{i=0}^1 \binom{1}{i} (8-1)^i = 8$$

which is equal to the number of states of Viterbi decoding, i.e., $q^{mk} = 8$.

The "pruned" error-trellis is shown in Fig. 1. And the construction of the trellis is described in the following. The labels on the pruned error-trellis shown in Fig. 1 correspond to the solution, Eq. (15a), of the syndrome equation for the actual error equal to the all-zero sequence. That is,

$$\begin{aligned} e &= [e_1, e_2] = uG = u[1 + D, 1 + \alpha D] \\ &= \{u + Du, u + \alpha Du\} \end{aligned} \quad (30)$$

is the output of the trellis. For example, at frame time j and state 0 if $u = \alpha^4$, then $e = [\alpha^4 + 0, \alpha^4 + \alpha \cdot 0] = [\alpha^4, \alpha^4]$ is

the label on transition from state 0 to state α^4 . Such a transition represents an attempt to "cancel" a simple error in the error-trellis equation, Eq. (15a). If such an error does, in fact, occur at frame j , then no further errors are allowed to occur at frame $j + 1$. Thus a transition to other than state zero must be followed by a transition back to state 0 in the next frame as shown in Fig. 1.

Next suppose a transition to state α^4 occurs, i.e., $Du = \alpha^4$. Then since $u = 0$, the transition from state $Du = \alpha^4$ back to 0 is given, using Eq. (30) by $e = [0 + \alpha^4, 0 + \alpha \cdot \alpha^4] = [\alpha^4, \alpha^5]$. The remaining labels to the "pruned" error-trellis are obtained in a similar manner.

To illustrate error-trellis decoding of the dual-3 CC let the generating sequence be

$$X(D) = 1 + \alpha D \quad (31)$$

Then the code word sequence is by Eqs. (7) and (26b)

$$Y(D) = X(D)G(D) = [1 + \alpha^5 D + \alpha D^2, 1 + \alpha^2 D^2]$$

Next, let the actual error sequence be $e_a(D) = [D^2, \alpha]$ so that

$$Z(D) = Y(D) + e_a(D) = [1 + \alpha^5 D + \alpha^5 D^2, \alpha^5 + \alpha^2 D^2] \quad (32)$$

Hence by Eq. (28),

$$\begin{aligned} ZR &= [1 + \alpha^5 D + \alpha^5 D^2, \alpha^5 D^2] \begin{bmatrix} \alpha^2 + \alpha^3 D, \alpha^3 + \alpha^4 D \\ \alpha^2 + \alpha^2 D, \alpha^3 + \alpha^3 D \end{bmatrix} \\ &= [\alpha^3, \alpha^4] + [\alpha^3, \alpha^4] D + [\alpha^2, \alpha^3] D + [\alpha^3, \alpha^4] D^3 \end{aligned} \quad (33)$$

The finding of the minimum weight error-path $\hat{e}(D)$ in terms of $u(D)$ as given by Eq. (17) is equivalent by Eq. (15a) to finding the code word $u(D)G(D)$ which is closest to $Z(D) \cdot R(D)$ as given in Eq. (33). Hence the minimum-weight error-path can be found by applying the Viterbi decoding algorithm (Ref. 3) to the pruned error-trellis in Fig. 1. To accomplish this, the frames of ZR in Eq. (33) are added to the output uG in the pruned error-trellis in Fig. 1 as shown in Fig. 2.

In order to illustrate the Viterbi algorithm as applied to the pruned error-trellis suppose the decoder has reached frame 4. The output of the transition from state α^3 to 0 is

$$\begin{aligned} \text{Coef } [u(D) \cdot G(D) + Z(D)R(D)] &= [\alpha^3, \alpha^4] + [\alpha^3, \alpha^4] \\ D^3 &= [0, 0] \end{aligned}$$

with Hamming weight 0. A similar calculation for the other 7 possible transitions shows that the transition from α^3 to 0 is the only one with Hamming weight 0. The path segment from α^3 to 0 is chosen since it has minimum weight.

At frame 5, Fig. 2, the minimum weight estimate of the D -transform of the error sequence is $\hat{e}(D) = [0, \alpha] + [1, 0] D^2$. Hence the estimate $u(D)$ of the message correction factor which achieves $e(D)$ is

$$\hat{u} = \alpha^3 + \alpha^3 D^2 \quad (34)$$

Finally, using Eqs. (27), (32), (34) in Eq. (18) yields, by Table 1,

$$\begin{aligned} \hat{X} &= ZG^{-1} - \hat{u} \\ &= Z \begin{bmatrix} \alpha^3, \alpha^2 + \alpha^3 D \\ \alpha^2, \alpha^2 + \alpha^2 D \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \hat{u} \\ &= [1 + \alpha^5 D + D^2, \alpha^5 + \alpha^2 D^2] \begin{bmatrix} \alpha^3 \\ \alpha^2 \end{bmatrix} \\ &\quad - [\alpha^3 + \alpha^3 D^2] = 1 + \alpha D \end{aligned}$$

the original encoded message in Eq. (31).

V. Recursive Algorithm for Error-Trellis Syndrome Decoding of Convolutional Codes

For the dual-3 CC described in Sec. IV, error-trellis decoding of CC is used to correct one error in every ℓ frames. For this example a recursive algorithm is developed to determine the path with minimum Hamming weight for every ℓ frames. This recursive algorithm for the error-trellis syndrome decoding of dual-3, rate 1/2, one-error-correcting nonsystematic convolutional codes is described with a flowchart as shown in Fig. 3.

The recursive algorithm is illustrated in example 1 of the last section. In this example, $u = 1$ and $\ell = 2$. By Eq. (33)

$$ZR = ([\alpha^3, \alpha^4], [\alpha^3, \alpha^4], [\alpha^2, \alpha^3], [\alpha^3, \alpha^4], [0, 0]) \quad (35)$$

In Eq. (35), the first value of ZR is not equal to zero. Thus it is assumed that one error occurs in the first two frames. To find this error, the pruned error-trellis with a no-error output $u(D)G(D)$ in $\ell = 2$ frames is computed. This partial trellis is shown in Fig. 4. An error-trellis over these two frames is created by adding the vectors $[\alpha^3, \alpha^4]$ and $[\alpha^3, \alpha^4]$, successively, to all labels, in the pruned trellis in Fig. 4. This resulting error-trellis for these two frames is shown in Fig. 5.

In Fig. 5 one needs only to find the path with minimum weight which ends up in state $\sigma = 0$ at the end of $\ell = 2$ frames. Using Viterbi decoding the path with minimum weight is $([\alpha, 0], [0, 0])$.

Next, the input value following $[\alpha^3, \alpha^4]$ and $[\alpha^3, \alpha^4]$ again is not equal to zero. Again it is assumed that only one error occurs in the next two consecutive values of $[\alpha^2, \alpha^3]$ and $[\alpha^3, \alpha^4]$. These values are added again to the pruned trellis of two frames given in Fig. 4 for generating an error-trellis. Again using Viterbi decoding one finds the path with minimum weight to be $([1, 0], [0, 0])$. Finally, the input value following $[\alpha^2, \alpha^3]$ and $[\alpha^3, \alpha^4]$ is $[0, 0]$. Hence the estimated error for these two frame times is $[0, 0]$. Thus the overall path with minimum weight is

$$\hat{e} = ([0, \alpha], [0, 0], [1, 0], [0, 0], [0, 0])$$

As a consequence, the correction factor $\hat{u}(D)$ is $\hat{u}(D) = \alpha^3 + \alpha^2 D^2$. Hence, from Eq. (18), the estimated message is $\hat{X} = 1 + \alpha D$. The same procedure applies similarly to a systematic one-error-correcting Wyner-Ash CC presented in Ref. 5. In this code, $\ell = m + 1 = 2 + 1 = 3$.

VI. A VLSI Design for Error-Trellis Syndrome Decoding of Convolutional Codes

In this section, a VLSI architecture is developed for the recursive algorithm for error-trellis decoding of convolutional codes presented in Fig. 3. This VLSI processor for selecting the path with minimum Hamming weight is composed first of d basic cells, where d is the number of paths in the error-trellis. These d cells are followed then by a weight comparison circuit. A basic cell computes the path weight incrementally. That is, if symbol A is not equal to symbol B , then the weight of that particular path increases by one; otherwise, the weight remains unchanged. The VLSI architecture of the error-trellis syndrome decoder is illustrated in the following example.

The calculations used in the present example were given in the last two sections. The VLSI architecture for this convolu-

tional code is illustrated in Fig. 6. In this figure there are 8 basic cells corresponding to the eight possible paths in the error-trellis. The function of each basic cell is described as follows.

The i th basic cell corresponds to the i th path in 2 frame times in the error-trellis. Thus if A is not equal to B , then the weight of the i th path increases by one (otherwise it remains the same) where A is the input value and B is the precalculated value stored in the i th basic cell. In this example, A is $(ZR)_k$, for $(k = 1, 2, 3, 4)$, and B is $(u \cdot G)_j$, for $(j = 1, 2, 3, 4)$, where Z is the received code sequence, R is defined in Eq. (28), u is an arbitrary k -vector of elements in $F[D]$, and G is the generator matrix.

First the received code sequence Z is multiplied by matrix R and G^{-1} , as shown in Fig. 7, to form both the input sequence to the basic cells and the inverse of the received message, i.e., $Z \cdot G^{-1}$. The input sequence ZR is then sent to all the 8 basic cells as well as a zero detector simultaneously. The inverse of the received message ZG^{-1} is sent then to a delay line to wait for the completion of the set of operations needed to estimate the correction factor \hat{u} . Then $Z \cdot G^{-1}$ is added to \hat{u} to obtain the estimated message \hat{x} by Eq. (18). The purpose of this zero-detector is to check if the input vector ZR is zero or not. If the two components are zero, then all the outputs of the weight comparator, which are described in the following, are also equal to zero. This indicates that the estimated correction factor u is zero, i.e., no error has occurred in the received code sequence Z . If ZR is not equal to zero, then the two components of ZR , i.e., $[ZR_1, ZR_2]$, in the first time frame are sent to the equality check circuit sequentially as shown in Fig. 8. The TG_i 's (for $i = 1, 2, 3, 4$) shown in Fig. 8 are 3-bit registers. They are used to store the precalculated $u_j \cdot G$ values for the j th path. Since it requires only 2 frame times to choose a minimum Hamming weight path, four registers are needed to store $u_j \cdot G$. The reason for the use of 4 registers instead of one in this design is to avoid a more complex sequential computation of $u_j \cdot G$.

The loading of the TG_i 's into the equality check circuit is controlled by a 2-bit counter which is capable of generating the required 4 different states. Because only 4 pairs of values need to be checked in every 2 frame times, four clock cycles are needed to finish the loading operation.

At the first clock cycle, ZR_1 and TG_1 are loaded into the equality check circuit. At the next clock cycle, ZR_2 and TG_2 are fed into the same circuit in sequence. The equality check circuit is implemented by the XOR arrays and an OR gate as shown in Fig. 9. For example, if ZR_1 is equal to TG_1 ,

then the output of the equality check circuit has logic level zero; otherwise it is one.

The output of the equality check circuit is sent to a 3-bit counter which accumulates the weight of a path in the error-trellis. After 4 clock-time, all the 4 pairs of the input sequence $Z \cdot R$ and $u_j \cdot G$ of each path on the error-trellis are compared. The output of the 3-bit counter which is the calculated weight of each path on the error-trellis is then sent to a weight comparator circuit. The weight comparator consists of Programmable Logic Array (PLA array). This is denoted by PLA1 in Fig. 10.

Also shown in Fig. 10 is an array of control gates and a table-lookup PLA. The inputs to each PLA1 in Fig. 10 are two 3-bit registers, W_i and W_j , which denote the weights of i th and j th path, respectively, in the error-trellis. The outputs of PLA1 are W_i or W_j depending on which is smaller, and a 1-bit signal, LR , to indicate if W_i is smaller than W_j . If W_i is smaller than W_j , then LR is zero; otherwise LR is one. The

PLA1's are configured in a tree structure. The outputs of the first level PLA1's are sent to the second level PLA1's as their inputs and so forth.

For example, if W_1 is the smallest of all the weights, as indicated in Fig. 10 at point A , it is logic zero. This will turn on gate T_1 and turn off gate T_2 . Then the value of C which is zero will be transferred through gate T_1 to point B . Since W_1 is the smallest value, the value at point D is zero. This turns on gates T_3 and T_4 and turns off gates T_5 and T_6 . Therefore the values at points A and B , which are zero, together with the value at D which is zero as well, are transferred to the inputs of another type of PLA, labeled by PLA2 in Fig. 10.

The function of PLA2 is to form a mapping between the path and correction factor \hat{u} . Since there are eight different paths in the error-trellis, there are, correspondingly eight different \hat{u} 's. Finally, the correction factor \hat{u} is added back to ZG^{-1} , by Eq. (18), as the estimated information \hat{X} . The estimated information \hat{X} is then shifted out of this circuit sequentially.

References

1. Reed, I. S., and Truong, T. K., "Error-Trellis Decoding of Convolutional Code" to be published in *Proceedings IEE*.
2. Reed, I. S., and Truong, T. K., "New Syndrome Decoding for $(n, 1)$ Convolutional Codes," *Electronic Letters*, Vol. 19, No. 9, April 1983, pp. 344-346.
3. Viterbi, A. J., and Omura, J. K., *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
4. Blahut, R. E., *Theory of Practice of Error Control Codes*. New York: Addison-Wesley, 1983.
5. Wyner, A. D., and Ash, R. B., "Analysis of Recurrent Codes," *IEEE Trans. Info. Theor.* IT-9, pp. 143-156, 1963.
6. Jensen, J. M., and Reed, I. S., "Error-Trellis Decoding of Convolutional Codes," submitted to *IEEE Trans. on Information Theory*.
7. Odenwalder, O. P., "Dual- K Convolutional Codes for Noncoherent Demodulated Channels," *Proceedings of International Telemetry Conference (ITC)*, Vol. 12, pp. 165-174, 1978.

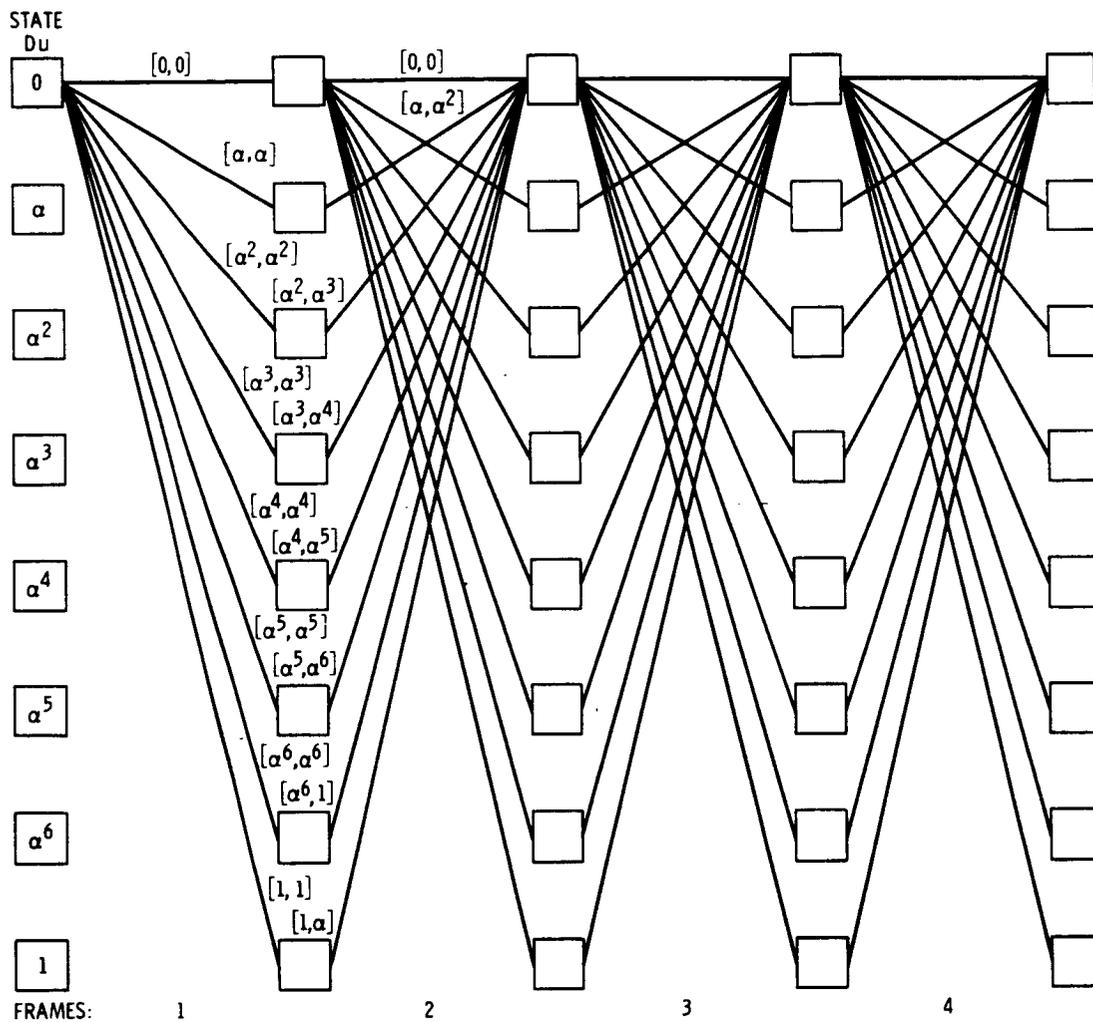


Fig. 1. Pruned error-trellis with no error outputs $u(D) G(D)$

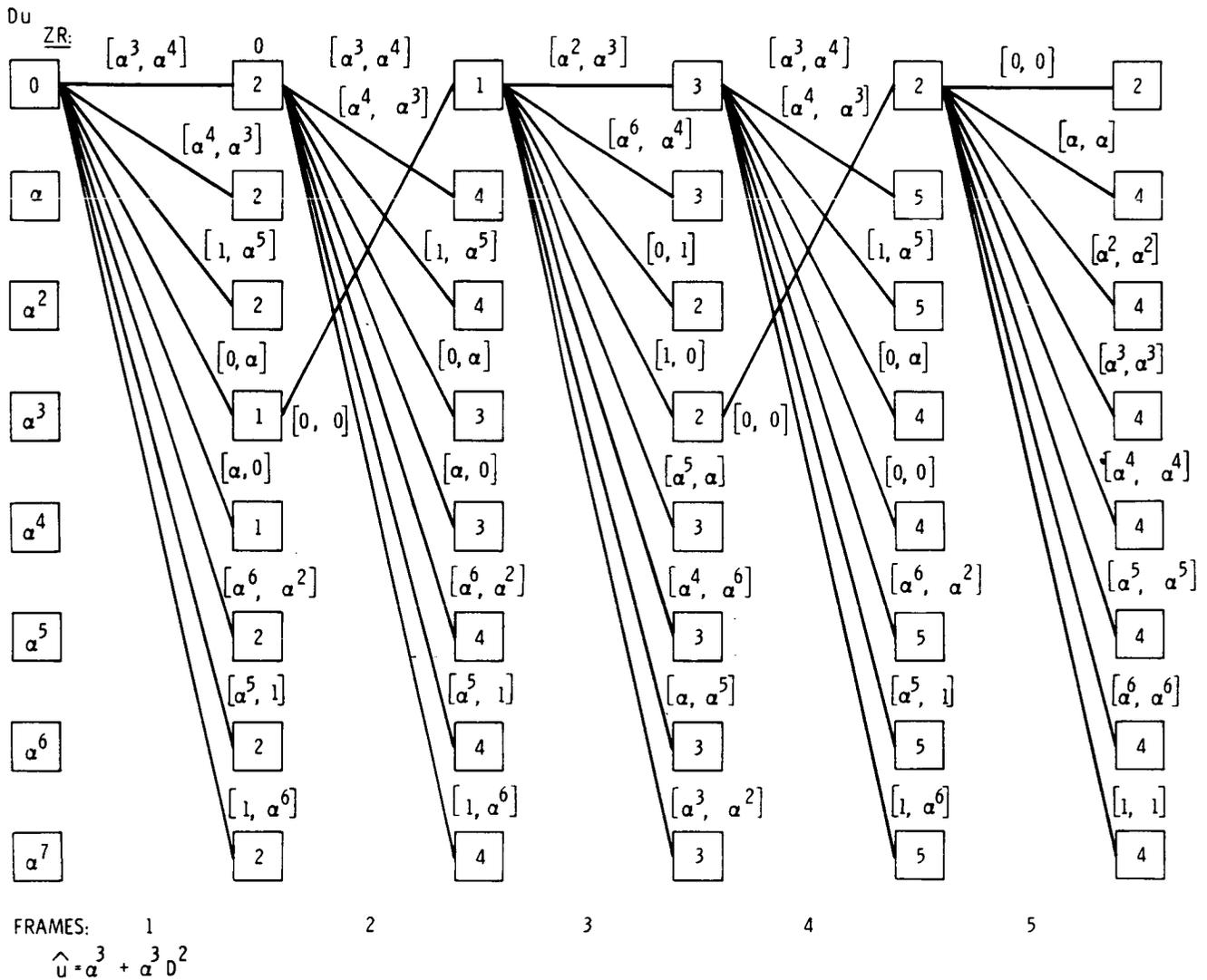


Fig. 2. Minimum-error path \hat{u} in pruned error trellis

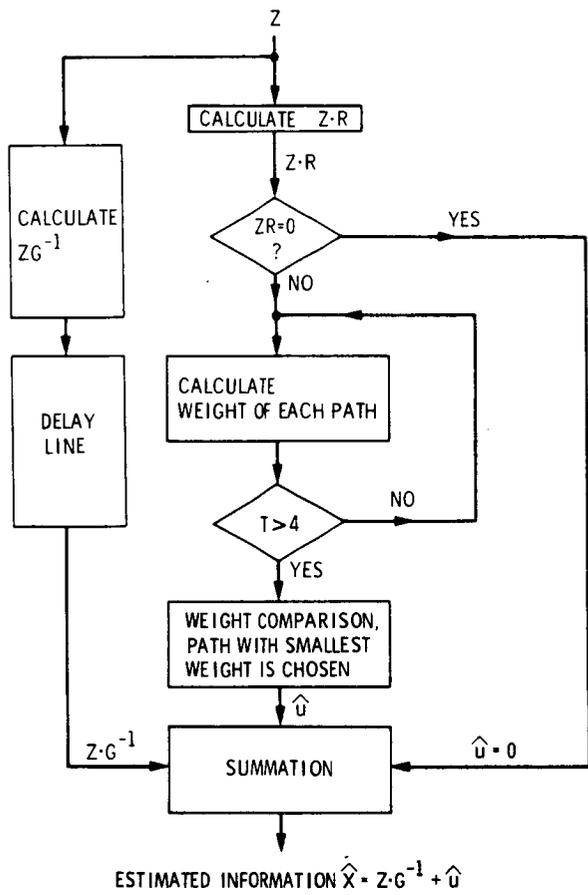


Fig. 3. Flow chart of error-trellis syndrome decoding of dual-3, rate 1/2, convolutional code

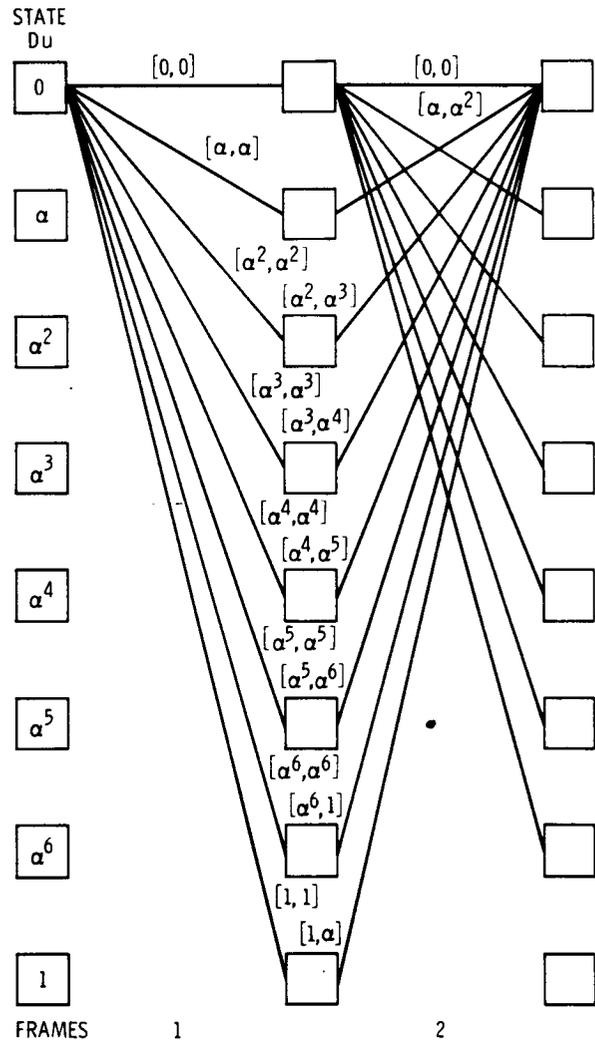


Fig. 4. A basic cell of a pruned error-trellis with no error outputs $u(D)G(D)$

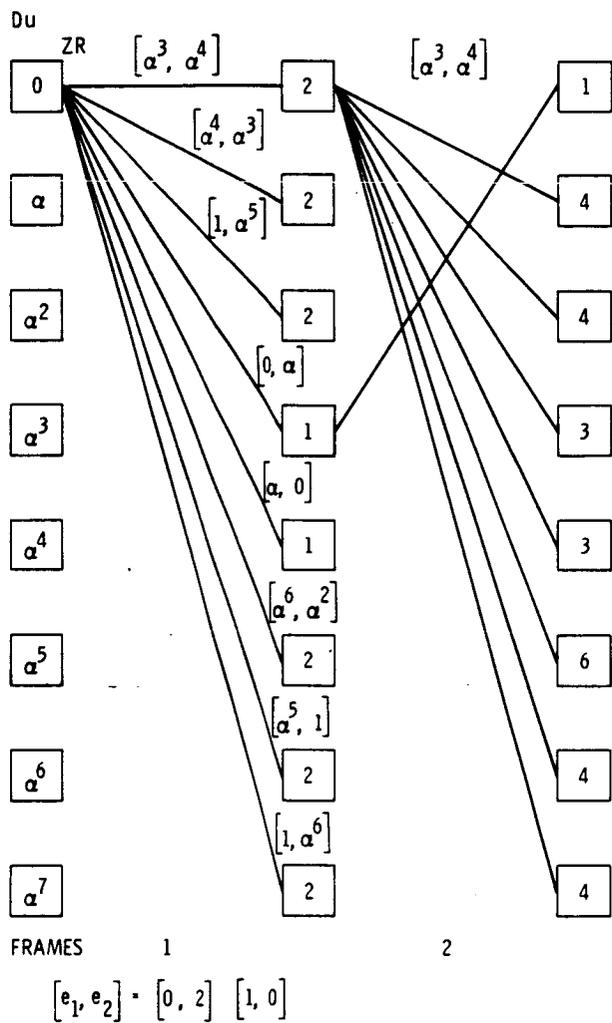


Fig. 5. Minimum-error path \hat{e} in pruned error trellis in two frames

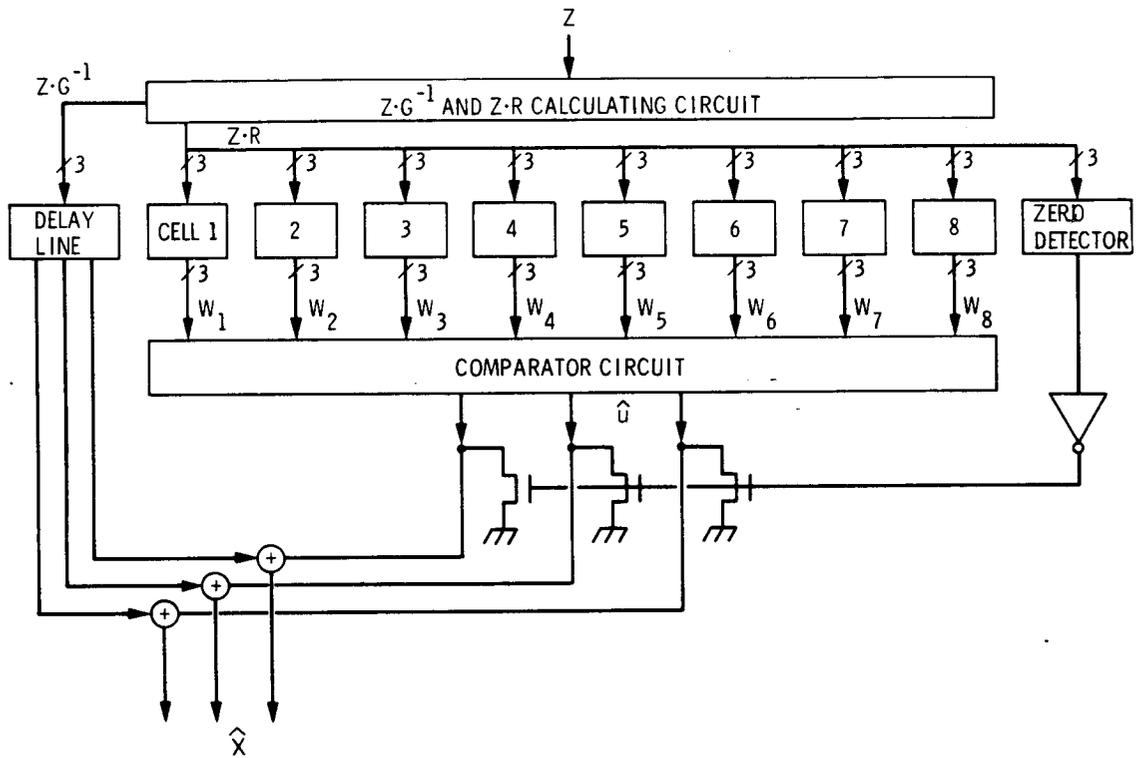


Fig. 6. Block diagram of a dual-3, rate 1/2, convolutional code decoder

$$ZR_1 = Z_1 (\alpha^2 + \alpha^3 D) + Z_2 (\alpha^2 + \alpha^2 D)$$

$$ZR_2 = Z_1 (\alpha^3 + \alpha^4 D) + Z_2 (\alpha^3 + \alpha^3 D)$$

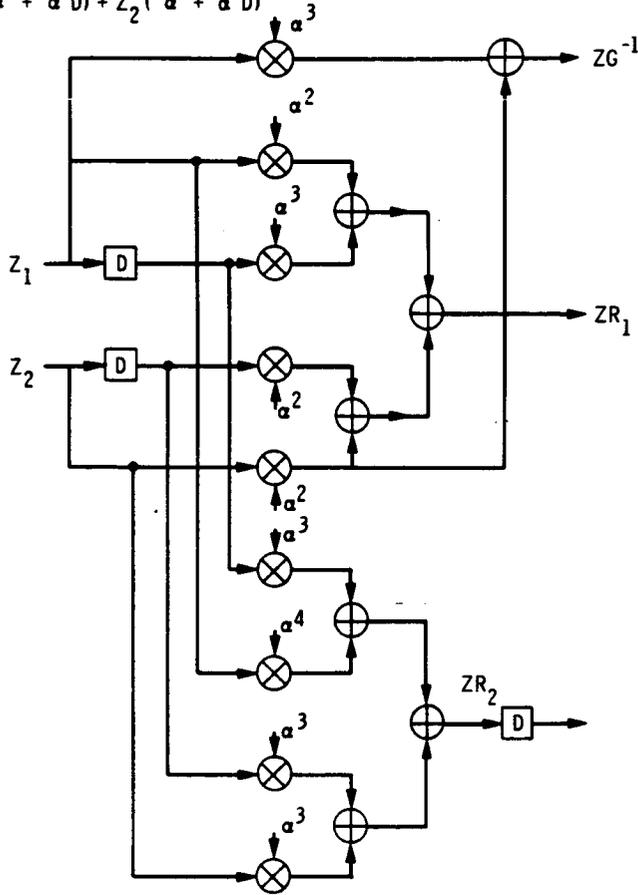


Fig. 7. Block diagram of circuit for calculating $Z \cdot R$ and $Z \cdot G^{-1}$

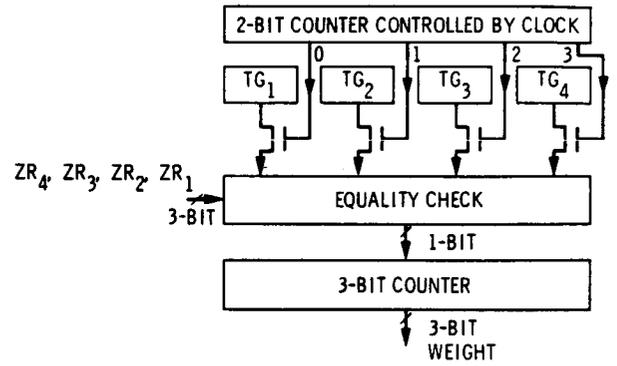


Fig. 8. Block diagram of each cell for calculating weight

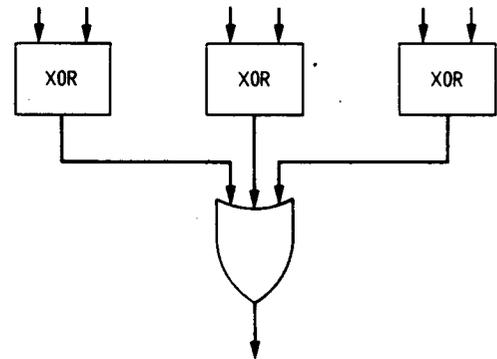


Fig. 9. Logic diagram of equality check circuit

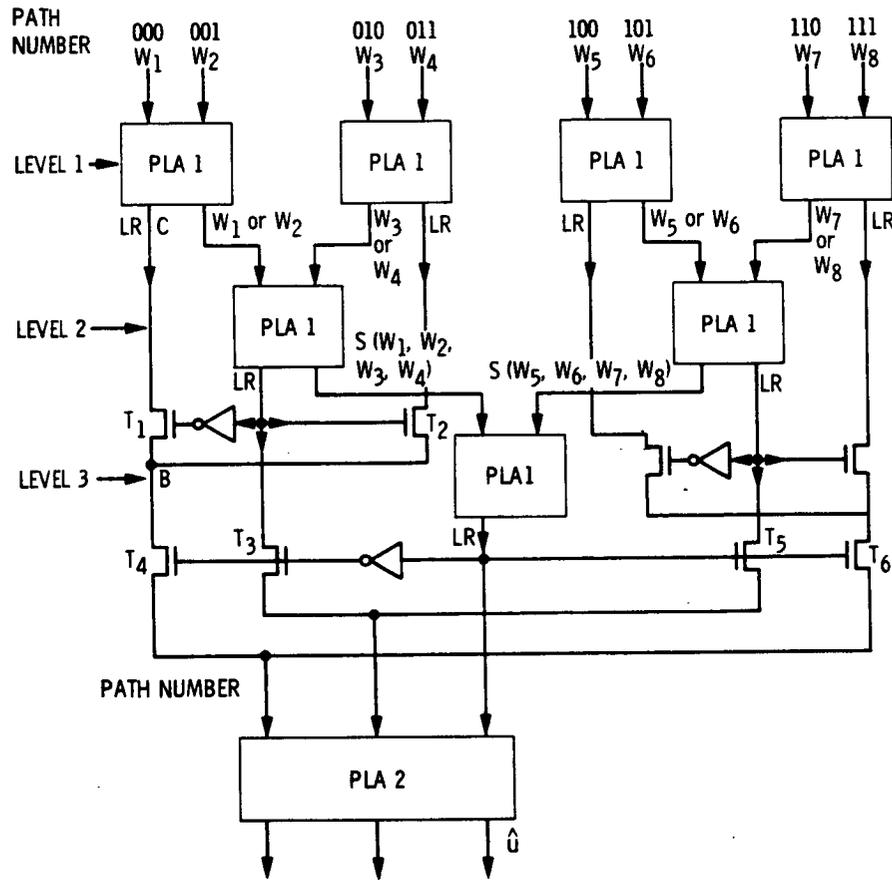


Fig. 10. Block diagram of weight comparator circuit. The inputs to this circuit are weights of each path. The outputs are estimated corrector factors.

Appendix A

Approximation of the Set $E^{(-1)}$ for Dual-K CCs

Let $G^{-1}(D)$ be a right inverse to the generator matrix (25b) of a dual-K CC of length n . Then $E^{(-1)} = \{eG^{-1} : eeE\}$. Define

$$\tilde{E}_1^{(-1)} = \{u : w_H(u_j, u_{j+1}) \leq t, \text{ for all } j \geq 0\} \quad (\text{A-1})$$

In the following it will be shown that $\tilde{E}_1^{(-1)}$ is a suitable approximation to $E^{(-1)}$. Since for $t \geq 2$, $E^{(-1)} \subseteq \tilde{E}_1^{(-1)}$, the only nontrivial case is for $t = 1$, that is for $n = 2$. It is readily verified that

$$G^{-1} = \left[\begin{array}{c} g_{12} \\ (g_{12} + g_{11}) \end{array}, \begin{array}{c} g_{11} \\ (g_{12} + g_{11}) \end{array} \right]^T = [g_1, g_2]^T \quad (\text{A-2})$$

is a right inverse matrix to the generator matrix of a dual-K CC of length 2. Note that g_1 and g_2 are nonzero elements in $GF(2^K)$.

Since the dual-K CCs of length 2 are 1-error-per-block-length-correcting CCs, the set of sequences

$$\tilde{E} = \{v : w_H(v_j, v_{j+1}) \leq 1, \text{ for all } j \geq 0\} \quad (\text{A-3})$$

is a desirable approximation to E . Thus

$$\tilde{E}^{(-1)} = \{v = vG^{-1} : v \in \tilde{E}\} \quad (\text{A-4})$$

is an approximation to $E^{(-1)}$. Also from the following lemma, one has that $\tilde{E}^{(-1)} = \tilde{E}_1^{(-1)}$.

Lemma. Let $n = 2$, then $\tilde{E}^{(-1)} = \tilde{E}_1^{(-1)}$.

Proof: Let $u(D) \in \tilde{E}^{(-1)}$. Then

$$u(D) = \sum_{j=0}^{\infty} u_j \cdot D^j = v(D) \cdot G^{-1}$$

where $v(D) \in \tilde{E}$. And from (A-2)

$$\begin{aligned} u(D) &= \sum_{j=0}^{\infty} (v_{1j} - v_{2j}) D^j [g_1, g_2]^T \\ &= \sum_{j=0}^{\infty} (v_{1j} g_1 + v_{2j} g_2) D^j \end{aligned}$$

Thus,

$$\begin{aligned} w_H(u_j, u_{j+1}) &= w_H(v_{1j} g_1 + v_{2j} g_2, v_{1j+1} g_1 + v_{2j+1} g_2) \\ &\leq w_H(v_{1j} g_1, v_{1j+1} g_1, v_{2j} g_2, v_{2j+1} g_2) \\ &= w_H(v_{1j}, v_{1j+1}, v_{2j}, v_{2j+1}) \\ &= w_H(v_1, v_2) \leq 1 \end{aligned}$$

From (A-3), since $v(D) \in E_1$, this implies that $E^{(-1)} \subseteq E_1^{(-1)}$.

To show that $\tilde{E}_1^{(-1)} \subseteq \tilde{E}^{(-1)}$, let $u(D) \in \tilde{E}_1^{(-1)}$, construct the sequence $v(D) = (v_1(D), v_2(D)) = (g_1^{-1} u(D), 0)$. Since $g_1^{-1} \cdot u(D) \in \tilde{E}_1^{(-1)}$, it follows from (A-3) that $v(D)$ is in \tilde{E} . Hence $v(D) \cdot G^{-1} = v(D) \cdot [g_1, g_2]^T = u(D)$ is in $\tilde{E}^{(-1)}$.

High-Rate Convolutional Code Construction With the Minimum Required SNR Criterion

P. J. Lee

Communications Systems Research Section

New short constraint length, high-rate convolutional codes which minimize the required SNR are found and tabulated for rates 2/3, 3/4, and 4/5, and for constraint length K up to 10. When compared with previously reported codes, most of the new codes reduce the required SNR only slightly. However, there are some pairs of K and code rate for which the new codes require considerably less SNR. The most significant one is the new $K = 8$, rate 4/5 code which requires 1.25 dB less SNR than the known code with the same parameters, for a desired bit error rate of 10^{-6} .

I. Introduction

For a convolutional coding system employing a Viterbi decoder, the decoded bit error rate (BER) is well upper-bounded by the transfer function bound (Refs. 1; and 2, Chap. 4)

$$\text{BER} \leq c_0 \cdot \frac{\partial}{\partial Z} T(D, Z) \Big|_{D=D_0, Z=1} = c_0 \cdot \sum_{i=d_f}^{\infty} a_i D_0^i \quad (1)$$

where the coefficient c_0 and transfer function $T(D, Z)$ depend on the code and type of channel used. The quantity D_0 is the Bhattacharyya bound (Ref. 2, p. 63) which depends on the channel only, d_f is the free distance of the code, and a_i is the number of bit errors in all incorrect coded symbol sequences with Hamming distance i . For an additive white Gaussian noise channel with binary PSK signaling (BPSK/AWGN channel) without quantization, we have (Refs. 1; and 2, p. 248)

$$D_0 = \exp(-E_s/N_0)$$

$$c_0 = Q(\sqrt{2d_f E_s/N_0}) \exp(d_f E_s/N_0)$$

where N_0 is the one-sided noise power spectral density, E_s is the received signal energy per channel symbol, and

$$Q(w) = \int_w^{\infty} \exp(-t^2/2) dt / \sqrt{2\pi}$$

Many researchers have used the maximum d_f criterion, or the criterion of maximum d_f together with minimizing the first few a_i 's in Eq. (1) for determining the goodness of a code in their code search procedures. However, we have shown in Refs. 3 and 4 that, for low rate codes, these criteria do not necessarily lead to codes which minimize required signal-to-noise ratio (SNR) for a certain desired BER. Direct use of Eq. (1) for BER evaluation in the search procedure provides much better results.

This new minimum required SNR criterion is applied here to the searches for good high-rate codes, which are useful for systems with limited bandwidth. In the next sections, our notation is introduced and the code search procedure is briefly explained. Search results are then listed and discussed.

II. Preliminaries

Let m_0 , k_0 , and n_0 be the number of binary memory cells, inputs, and outputs of an $(m_0, k_0/n_0)$ convolutional encoder, where k_0/n_0 is the code rate r . A typical nonsystematic, time invariant encoder structure is shown in Fig. 1. A group of k_0 information bits is shifted into a shift register of length $K (= m_0 + k_0)$, and outputs of n_0 modulo-2 adders are sampled and sequentially transmitted. The parameter K is called the constraint length of the code, while m_0 is called its memory length. Notice that the number of states in the Viterbi decoder trellis is 2^{m_0} . The low-rate codes considered in Refs. 3 and 4 are special cases with $k_0 = 1$.

Besides these key parameters, the code performance is determined by the connections from K shift registers to n_0 modulo-2 adders. These connections are often represented by an $n_0 \times K$ binary matrix G , called the code generator matrix, where "1" stands for connection and "0" for non-connection. As an illustration, a $(2, 3/4)$ encoder and a $(3, 2/3)$ encoder are shown in Fig. 2, whose code generator matrices are given, respectively, by

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

and

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

For short hand notation, G is often represented by

$$(g(1), \dots, g(j), \dots, g(n_0))$$

where $g(j)$ is the j th row of G , in octal. For the codes in Fig. 2, $G = (37,21,5,4)$ and $G = (31,23,16)$, respectively. By "code search" we imply the search for a code generator G which provides good performance among the codes with same key parameters.

The transfer function bounding technique on the BER at the Viterbi decoder output will not be discussed here; it can be found in many references, including Refs. 5, 6, and 7.

III. Code Searching Procedure

In this study, we restricted our searches to high-rate codes with $n_0 = k_0 + 1$. Notice that the number of $(m_0, k_0/(k_0 + 1))$ codes in the whole code space is $2^{(m_0+k_0) \times (k_0+1)}$. For example, there are over 4 billion $(5, 3/4)$ codes. Furthermore, according to our criterion, to test a code we have to evaluate the transfer function bound, which requires a matrix inversion. Therefore an exhaustive search is prohibitively difficult except for very small m_0 and k_0 . Only partial searches are possible to obtain results in a reasonable length of time.

In the previous searches for low-rate codes, we developed several effective techniques for reducing the code search space. Many of these techniques are applied to the high-rate code searches with appropriate modifications.

First, for a given pair of m_0 and k_0 , we made a list of some $r = 1$ codes (actually these are not codes since there is no redundancy) which are to be used for the generation of $r = k_0/(k_0 + 1)$ codes. In this list, by using the simple fact that exchanges of $g(j)$'s do not affect the performance of the code, identical codes are discarded. Also, codes with too small free distance (less than $d_x - 2$, where d_x is the maximum known free distance of $(m_0, k_0/(k_0 + 1))$ codes, or its bound if not known) are deleted. This procedure is based on the observation (Ref. 3) that the Hamming distance (from the all-zero output) on each branch (in the state diagram) of a lower rate code is always larger than or equal to that of the higher rate code, used as a seed for its generation. Catastrophic codes are not discarded at this time, as good $r = k_0/(k_0 + 1)$ codes are often found from catastrophic $r = k_0/k_0$ codes.

Each code in the list is used for generation of lower rate codes. Among the generated codes, identical codes and catastrophic codes are deleted. Codes with free distance smaller than $d_x - 1$ are also discarded. For each remaining code, the BER performance is found by the transfer function bound with a SNR at which the best code is expected to achieve BER of 10^{-6} .

IV. Search Results

The code search results are summarized in Table 1, where the code generators of best codes are shown with their free distance and the upper bound on the required bit SNR

$(E_b/N_0, E_b = E_s/r = E_s \cdot (k_0 + 1)/k_0)$ value for desired BER of 10^{-6} . These new codes are compared with the codes reported in Refs. 8 and 9. If the best code in the sense of minimum required SNR does not have maximum free distance, then the best code among maximum free distance codes is also listed.

Notice that some of new codes have parameters never considered before. For all codes, we were able to find better

codes than the previously reported codes. But the amount of SNR saving is usually very small except for a few cases. For the case of (4, 4/5) code (or equivalently $K = 8, r = 4/5$ code), we found a code which not only requires 1.25 dB less SNR but also has larger free distance than the previously reported code. Since we could not exhaust the code search space for most cases, there might be some better codes. However, we expect that better codes, if they exist, would improve the performance very slightly.

References

1. Viterbi, A. J., "Convolutional codes and their performance in communications systems," *IEEE Trans. Commun. Tech. COM-19*, pp. 751-772, Oct. 1971.
2. Viterbi, A. J., and Omura, J. K., *Principles of Digital Communication and Coding*, New York: McGraw-Hill Book Co., 1979.
3. Lee, P. J., "New short constraint, rate $1/N$ convolutional codes which minimize the required SNR for given desired bit error rate," *IEEE Trans. Commun. COM-33*, pp. 171-177, Feb. 1985.
4. Lee, P. J., "Further results on rate $1/N$ convolutional code construction with minimum required SNR criterion," *TDA Progress Report 42-80*, pp. 97-102, Jet Propulsion Laboratory, Pasadena, CA., Feb. 1985.
5. Omura, J. K., and Simon, M. K., *Modulation/Demodulation Techniques for Satellite Communications, Part IV: Appendices, JPL Publication 81-73*, Jet Propulsion Laboratory, Pasadena, CA., Nov. 1981.
6. Lee, P. J., "Transfer function bound for partial-unit-memory convolutional codes based on reduced state diagram," *TDA Progress Report 42-76*, pp. 70-76, Jet Propulsion Laboratory, Pasadena, CA., Feb. 1984.
7. Lee, P. J., "A very efficient transfer function bounding technique on bit error rate for Viterbi decoder, rate $1/N$ convolutional codes," *TDA Progress Report 42-79*, pp. 114-123, Jet Propulsion Laboratory, Pasadena, CA., Nov. 1984.
8. Paaske, E., "Short binary convolutional codes with maximal free distance for rate $2/3$ and $3/4$," *IEEE Trans. Info. Theory IT-20*, pp. 683-689, Sept. 1974.
9. Daut, D. G., Modestino, J. W., and Wismer, L. D., "New short constraint length convolutional code construction with selected rationale code rates," *IEEE Trans. Info. Theory IT-28*, pp. 794-800, Sept. 1982.

Table 1. Best $(m_0, k_0/n_0)$ binary convolutional codes with $n_0 = k_0 + 1$ which minimize the required SNR for BER $\leq 10^{-6}$ and performance comparison to previously reported codes

$(m_0, k_0/n_0)$	d_f	E_b/N_0 , dB	Note	Code generator G, in octal				
(1, 2/3)	2	9.023	A	5	3	2		
	2	9.032	D	6	3	2		
(2, 2/3)	3	7.360	A	15	13	12		
	3	7.570	D	17	15	6		
(3, 2/3)	3	7.798	P	16	13	11		
	4	6.292	A	31	23	16		
	4	6.320	D	33	22	15		
(4, 2/3)	4	6.341	P	37	22	11		
	5	5.870	A	61	46	37		
	5	5.888	P	61	56	27		
(5, 2/3)	5	6.169	D	75	72	27		
	6	5.531	A	171	112	73		
	6	5.580	P	177	112	55		
(6, 2/3)	6	5.171	A	366	241	163		
	7	5.211	P	337	236	155		
(7, 2/3)	7	4.846	A	751	522	343		
	8	4.853	A	673	465	262		
(8, 2/3)	8	4.883	P	751	532	367		
	8	4.632	A	1671	1322	423		
(1, 3/4)	2	8.633	A	15	12	4	2	
	2	8.639	D	15	14	13	2	
(2, 3/4)	3	7.527	A	37	21	5	4	
	3	7.634	D	36	32	14	7	
(3, 3/4)	4	6.629	A	67	51	43	25	
	4	6.652	D	61	47	25	13	
(4, 3/4)	4	6.042	A	157	122	41	24	
	4	6.336	D	172	127	106	45	
(5, 3/4)	5	5.735	A	255	236	164	127	
	5	5.776	D	357	216	124	45	
	5	5.797	P	367	244	141	72	
(6, 3/4)	6	5.449	A	723	657	345	261	
	6	5.452	P	512	467	311	274	
(7, 3/4)	6	5.185	A	1752	1233	756	377	
(1, 4/5)	2	8.578	A	34	23	10	4	2
	2	8.820	D	36	26	13	11	4
(2, 4/5)	2	8.003	A	71	53	34	10	4
	2	8.507	D	67	57	52	26	15
(3, 4/5)	3	6.760	A	153	137	51	25	15
	3	6.838	D	174	132	56	23	13
(4, 4/5)	4	6.316	A	373	254	225	215	112
	3	7.561	D	337	274	255	237	156
(5, 4/5)	4	5.993	A	765	613	571	537	110
(6, 4/5)	5	5.710	A	1537	1351	1145	1053	730

NOTES: A Found by the author
P Found by Paaske (Ref. 8)
D Found by Daut, et al. (Ref. 9)

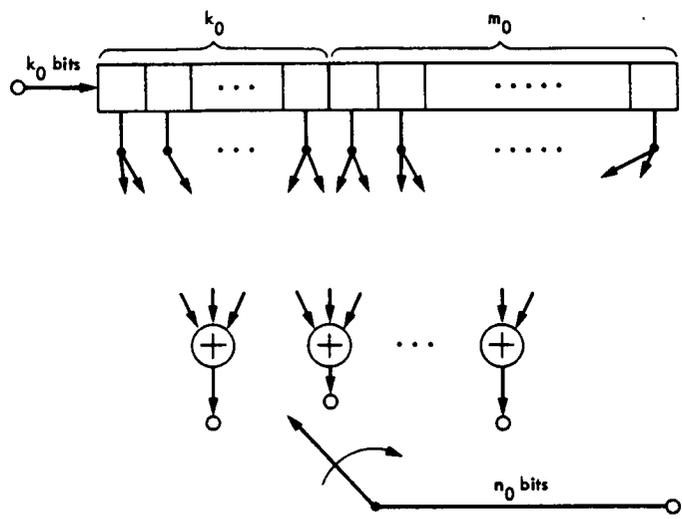


Fig. 1. A typical $(m_0, k_0/n_0)$ encoder structure

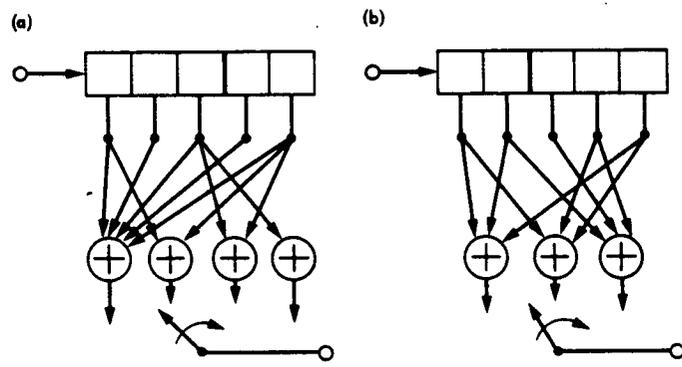


Fig. 2. Examples: (a) A $(2, 3/4)$ code and (b) a $(3, 2/3)$ code

DSN Frequency and Timing System, Mark IV-85

B. W. Falin

TDA Engineering Section

As part of the DSN Mark IVA Implementation Program, the DSN Frequency and Timing System is currently undergoing changes. With the implementation of Signal Processing Centers (SPC) at each Deep Space Communications Complex (DSCC), major changes to the frequency and timing distribution equipment were necessary. This article provides a functional description of the Mark IVA Frequency and Timing System (FTS) as it exists today and planned capabilities through 1988.

I. System Definition

A. Background

The necessity to upgrade the Mark III-77 FTS was driven by two factors. First, the consolidation of the DSN stations at a central complex required an increased distribution capability for time and reference frequencies. It also required centralized control and monitor of the frequency and timing equipment. Since a complex would be supporting concurrent operational activities, a need for the generation and distribution of simulated time was necessary. The second factor was the growing need to replace aging equipment that had been in service for twenty years.

B. Functional Definition

The DSN Frequency and Timing System provides precision frequency and timing data required at the Deep Space Stations throughout the DSN.

The DSN Frequency and Timing System performs two main functions, providing frequency and timing data and validation. Frequency and timing data consists of sinusoidal

reference frequencies, timing pulses, and epoch time codes. Frequency and timing validation assures that the primary frequency standards throughout the entire DSN are maintained within prescribed tolerances relative to the National Standard-United States Naval Observatory/National Bureau of Standards (USNO/NBS).

C. Functional Description

The Frequency and Timing System functions are conceptualized as follows:

- (1) Generation of precision frequency signals to provide references for the generators of sinusoidal signals, timing pulses, and time codes within DSCC's.
- (2) Generation and distribution of sinusoidal reference frequency signals, timing pulses, time codes, and provision for time displays within DSCC's.
- (3) Control and monitoring of configuration status and operating mode of the FTS at DSCC's.
- (4) Network-wide synchronization of all Frequency and Timing System signals to NBS, validation of frequency

standard calibration, and traceability of time to Universal Time, Coordinated (UTC).

Figure 1 depicts the Mark IVA Frequency and Timing System configuration.

D. Key Characteristics

The key characteristics of the Frequency and Timing System are:

Knowledge of time synchronization (to $\pm 10 \mu\text{s}$) and frequency offset (± 3 parts in 10^{-13}) between DSCC's

Knowledge of time synchronization (to $\pm 5 \mu\text{s}$) between the DSN and the National Standard (USNO/NBS)

Knowledge of time synchronization (to $\pm 10 \mu\text{s}$) and frequency offset (to ± 1 part in 10^{-11}) within each DSCC

Performance validation of frequency and timing functions for each Front End Area (FEA) and SPC at a DSCC

Maintenance of a permanent record of frequency and time parameters including configuration and synchronization

Simulation time selectable from the DSCC Monitor and Control Subsystem (DMC).

II. System Elements

A. Functional Elements

The major elements of the Frequency and Timing System consist of a Frequency and Timing Subsystem located at each DSCC and a Network FTS Data Base subsystem located at DSN Network Operations Control Center. The Frequency and Timing Subsystem consists of six major subsets of equipment. (The relationships between the subsets are depicted in Fig. 2.)

- Frequency reference standards
- Frequency synthesis and distribution
- Reference frequency cable stabilization
- Time and timing generation and distribution
- Time synchronization
- Monitor and control.

B. DSS Frequency and Timing Subsystem

The master frequency standard consists of two Hydrogen Masers backed by two Cesium Beam Frequency Standards.

These standards, which are located in an environmentally controlled area, are provided with an uninterruptible power source for emergencies. Any of the four standards can be selected by the complex operator as the prime standard. However, if the prime standard fails, another standard is automatically selected.

The frequency synthesis and distribution subset consists of a Coherent Reference Generator (CRG), which includes discrete synthesis, distribution amplifiers, and an interface panel. The CRG receives 0.1 MHz, 1 MHz, 5 MHz, and 100 MHz from the on-time frequency standard. It generates 10.1 MHz, 45 MHz, 50 MHz, and 55 MHz and provides distribution ports for user subsystems.

Sinusoidal reference signals are also distributed to users at the antenna areas. This is accomplished by the use of active stabilized cables that compensate for small phase changes in the distribution path. These actively stabilized cables are used for VLBI to provide the same stability as that of the Hydrogen Maser reference frequency standard. The stability requirements for the standard reference frequency are defined in Table 1.

The time and timing subset consists of a triple redundant master clock, simulation time generator, time insertion and distribution, and time code translators. The master clock equipment provides the following new capabilities for the DSN:

- Year End—Automatic reset
- Leap Year—Automatic extra day addition
- Leap Second—Automatic subtraction or addition of leap second
- Resetability—Simple clock adjustments

Time codes and timing pulses are provided to each user via a time code translator. Two types of translators are provided for user application, and the time code is available in binary milliseconds or binary-coded decimal (BCD) seconds. Time codes are also distributed to users at the antenna areas via fiber optic cables.

Time synchronization of the three complexes to the National Bureau of Standards is accomplished via the Global Positioning Satellite (GPS) System by utilizing a GPS receiver at each complex. Also, other sources for time synchronization include: WWV, LORAN C, TV line 10, and traveling clocks. Time-sync data are also available from the VLBI system.

The functional capabilities of the Frequency and Timing Subsystem can be controlled locally at the equipment level or

remotely from the complex operator position. The complex operator is provided with various displays that contain both status and performance parameters. Such parameters will be forwarded to NOCC for application in the Network FTS Data Base.

C. NOCC Subsystem

The Network FTS Data-Base Subsystem, to be implemented in 1988 at NOCC, will receive configuration data, status data, and performance data from the FTS subsystem at each DSCC. Also it will receive GPS time-sync data from the Tracking System and VLBI time-sync data from the VLBI System. The data will be analyzed as to time and frequency offsets between complexes, clock behavior, and Hydrogen Maser performance. The results will be available to complexes and to Network Operations.

Presently, the analysis function is being performed by the Network Operations and Analysis Section (NOA) as a DSN supporting element. NOA is responsible for generating reports,

monitoring the FTS system, and coordinating changes to complex time and frequency standards, as well as generating monthly reports containing both time and frequency measurement data.

III. FTS System Schedule

The following additional implementations will complete the Mark IVA FTS:

In support of the X-band uplink implementation, stabilized cables and distribution equipment are required for distributing 100-MHz reference signals on the antennas at DSS 45, 65, and 15.

Time analyzer assemblies and frequency standard monitor equipment will be implemented as part of the FTS monitor and control in FY 86.

FTS Data-Base Subsystem at Network Operations Control Center will be implemented in FY 88.

Table 1. Reference Frequency Stability

<i>T</i>	Allen Variance (σ)	
	Signal Processing Center	VLBI Antenna-Mounted Equipment
1 sec	1×10^{-12}	1×10^{-12}
10^4 sec	1×10^{-14}	1×10^{-14}
12 hours	1×10^{-14}	1×10^{-14}
10 days	1×10^{-13}	1×10^{-13}

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

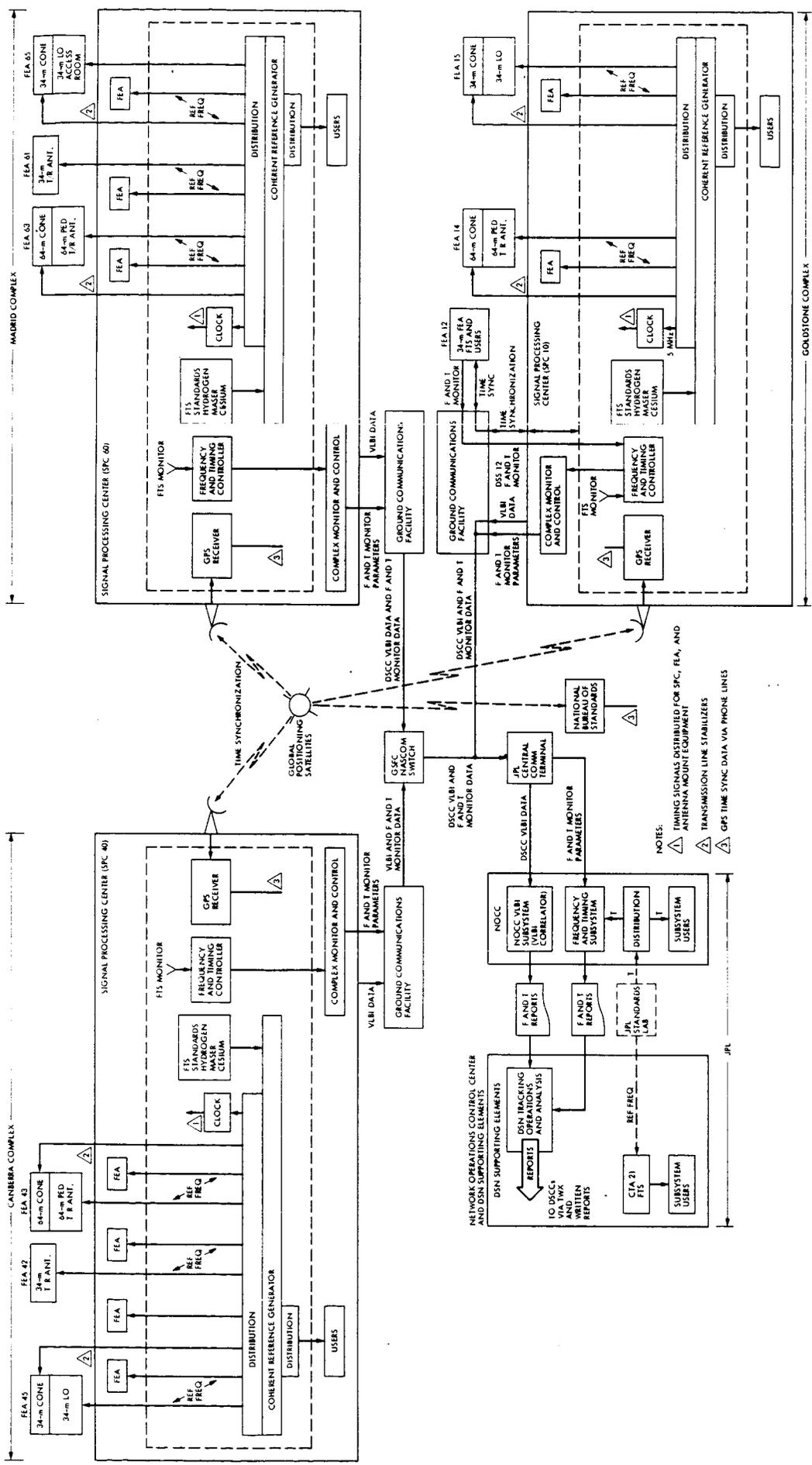


Fig. 1. Mark IVA DSN Frequency and Timing System functional design block diagram

FOLECUT FRAME

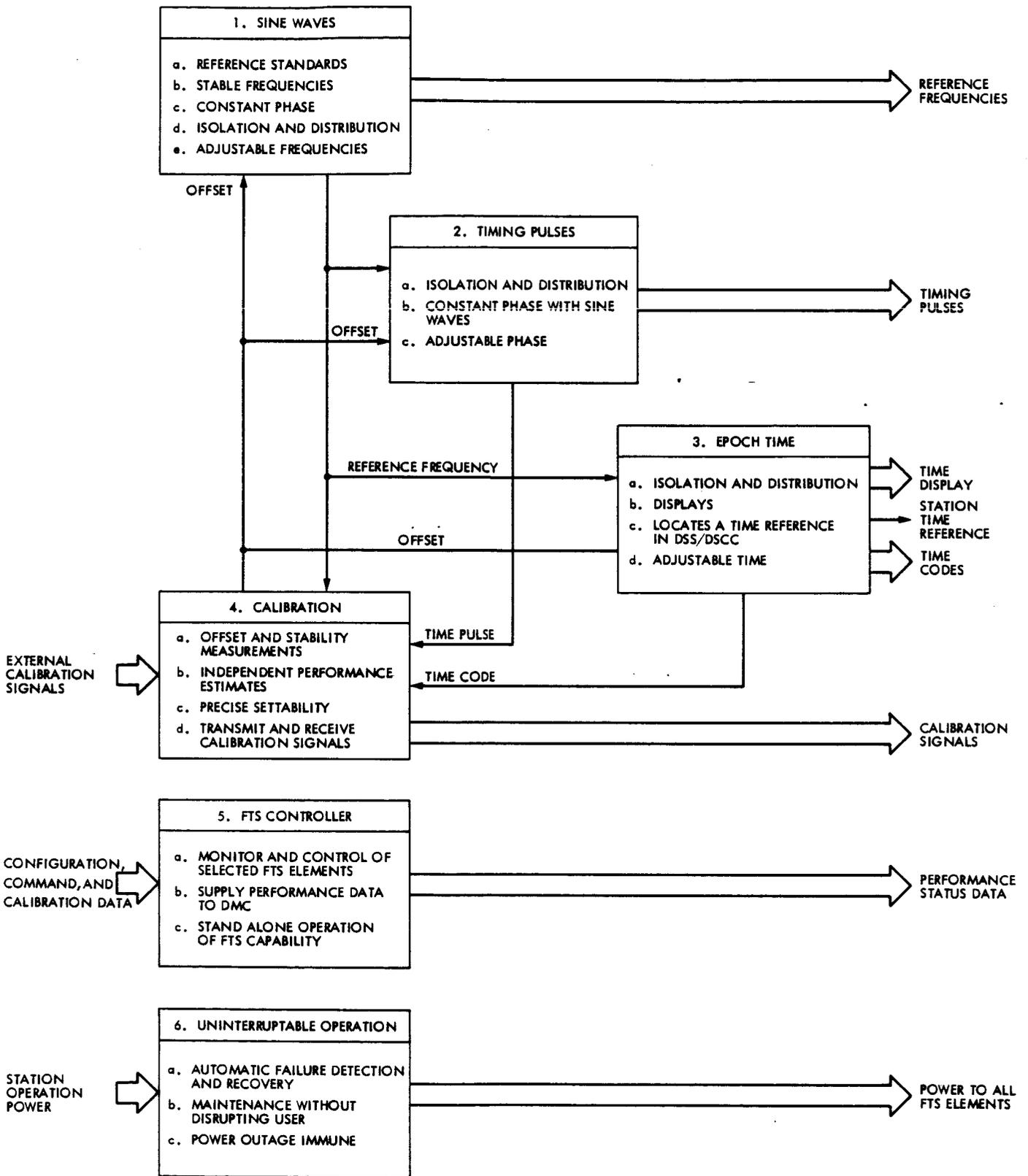


Fig. 2. FTS functional elements and relationships

JPL Emergency Support of TDRSS and Compatible Satellites

N. A. Fanelli

TDA Mission Support and DSN Operations Office

The Tracking and Data Relay Satellite System (TDRSS) will consist of three identical satellites in geosynchronous orbits and a dedicated ground receiving station. The first two satellites (TDRS East and TDRS West) will form the operational TDRS service network providing near-global real-time user satellite coverage. The third TDRS satellite will act as an in-orbit spare. Since the TDRSS satellites are supported by a single ground station, a method of providing emergency support for TDRS and user satellites is needed. This article describes the support to be provided by JPL's Deep Space Network.

I. Background

With the concept of an orbiting Tracking and Data Relay Satellite System (TDRSS), NASA's Office of Space Tracking and Data Systems (OSTDS) adopted a policy of discontinuing the support of low earth orbiting satellites with a ground based network of tracking stations.

Goddard Space Flight Centers (GSFC), Flight Projects Directorate, requested a reconsideration of this policy in 1983. The Jet Propulsion Laboratory (JPL) was requested to assess the impact of providing support for a limited number of projects under emergency conditions.

JPL's study showed that for a nominal cost, the 26-meter subnetwork of antennas could provide the requested support. OSTDS then approved the concept and assigned JPL the responsibility of providing this support.

II. Support Conditions

Low earth orbiting satellites will normally be supported by the TDRSS and the White Sands, New Mexico, receiving station. In the event of a failure, either ground station or satellite, which would prevent or limit the normal communications functions, JPL may be requested to provide emergency support. GSFC is responsible for the JPL notification that an emergency exists.

III. Supported Projects

The following satellites will be supported.

A. Solar Maximum Mission (SMM)

This satellite studies solar activity occurring in active regions of the sun, sunspots, and solar flares. Observations are

conducted in the ultraviolet, X-Ray, and Gamma-Ray regions of the spectrum with earth based instruments providing coverage at radio and optical wavelengths.

B. Land Satellite 4 and 5 (Landsat 4 and 5)

Landsat provides data continuity of earth resources information for worldwide users. Each satellite contains a thematic mapper and a multispectral scanner imaging device plus mission unique hardware.

C. Earth Radiation Budget Satellite (ERBS)

ERBS studies the earth's interaction with solar energy, a critical factor in predicting the earth's climate.

D. Space Telescope (ST)

The Space Telescope is planned as a national facility. It will consist of a 2.4-meter aperture Ritchey-Chretien cassegrainian telescope weighing approximately 9525 kilograms with various energy detectors designed for observations of infrared, visible, and ultraviolet wavelengths.

E. Gamma Ray Observatory (GRO)

GRO is an earth orbiting satellite that will study sources of localized, galactic, and extragalactic gamma radiation.

F. Upper Atmosphere Research Satellite (UARS)

UARS is designed to study the radiation, chemistry and dynamics of the upper atmosphere at low, middle, and moderately high altitudes and the coupling between these properties in order to determine the seasonal correlations.

G. Space Transportation System (STS)

The Space Transportation System is the manned reusable launch vehicle (Shuttle) used to carry into space nearly all of the nation's payloads for military, private industry, universities, research organizations, and foreign governments and organizations.

H. Tracking and Data Relay Satellite (TDRS)

The payload of each TDRS provides a Telecommunications service system which relays communications signals between low earth orbiting user spacecraft and the TDRS ground terminal. The service is provided by two types of links: (1) a multi-access system, with one 30-element S-Band phased array antenna system; and (2) a single-access system, either S-Band or K-Band, using two 4.8-meter parabolic antennas.

IV. Support Requirements

A. Initial Action

Upon initial notification of a declared satellite emergency requiring JPL support, JPL will take the steps necessary to provide station support. The DSN stations will acquire real-time engineering telemetry so that the project may assess the satellite health. They will also acquire high rate telemetry data recorded on the satellite's magnetic tape recorder. The recorder playback will be activated by ground command. The supporting DSN stations will generate and transmit radio metric data. All data will be provided to GSFC in real time or near real time.

B. Subsequent Passes

For subsequent passes, JPL will acquire real time engineering telemetry data, transmit ground commands to reload the spacecraft memory, if requested, acquire spacecraft memory dumps of telemetry data, acquire radio metric data, and provide data to GSFC in real time.

C. Prolonged Support

For prolonged emergency situations, reduced mission objectives will be pursued. Telemetry data will be acquired in real time, science data will be acquired via spacecraft recorder playbacks lasting from six to eight minutes, commands will be transmitted on a scheduled basis, radio metric data will be acquired, data rates exceeding the line capacity will be recorded on-station and played back during the post-pass period, and spacecraft coverage will be provided as defined in a negotiated schedule.

D. Support Configuration

TDRSS spacecraft emergency support will be provided by JPL's Deep Space Network (DSN) 26-meter subnetwork of stations located at Goldstone, California, Deep Space Station 16 (DSS 16); Canberra, Australia, (DSS 46); and Madrid, Spain, (DSS 66). The configuration delivered by the networks consolidation program will be used and will have the capability of supporting the data rates listed in Table 1. Real time transmissions will normally be limited to data rates which can be accommodated by a 56-kilobit transmission line rate. Under certain emergencies this capability will be increased to a transmission rate of 112 kilobits. Figure 1 shows a typical support configuration. Documentation will consist of a network operations plan (NOP) for each mission to be supported. This plan will document the plans, procedures, and configurations to be followed by the DSN stations, the DSN operations control team, and the DSN ground communications facility following the declaration of an emergency. The data interface for each mission to be supported will be documented in the

interface control document (ICD) for GSFC and JPL. The ICD will describe the operational, technical and communications aspects of the data transfer between centers and projects using the DSN.

E. Support Scenario

When an emergency situation develops, GSFC will notify the DSN operations chief using the coordination voice network. This voice network is active 24 hours per day, seven days per week and connects the GSFC network control center to the DSN network control center. GSFC will inform the DSN of the circumstances for declaring the emergency, the parameters needed to acquire the spacecraft such as downlink frequency or channel, telemetry data rate and coding, uplink frequency or channel, last received uplink and downlink signal levels, etc. Arrangements will also be made for the transmission of vectors needed to point the 26-meter antenna at the spacecraft.

The operations chief will assess the request and negotiate and schedule station support. A network controller will be assigned to monitor the station to project interface; when necessary, communications lines will be activated between the assigned station and the project's operations control center at GSFC. The station will be notified by the DSN controller of

the declared spacecraft emergency and will be asked to terminate the activity in progress and prepare for the emergency support situation. The controller will brief the station on the nature of the emergency and provide the parameters needed for spacecraft acquisition.

The GSFC Flight Dynamics Facility will provide state vectors to JPL or transmit an improved interrange vector (IIRV) directly to the station. These vectors will be used for antenna pointing. The station will use the NOP to configure the station for support. Following the completion of the preacquisition checklist the station will point the antenna at the spacecraft and acquire the downlink signal. At this point the station will notify the DSN controller of the acquisition of signal (AOS).

The project will instruct the DSN Controller on requirements for recordings, ranging, and command. During the pass, the DSN controller will monitor the net and provide assistance to the project or station. Upon completion of the pass, the station will be released for the next scheduled activity. The project will determine if continued emergency support is required and provide the DSN scheduling office with the requirements.

Table 1. Emergency satellite data rates

Mission	Initial R/T Telemetry, kbps	Contact Rates: Recorder Dumps, kbps	Subsequent R/T Telemetry, kbps	Contact Rates: Memory Dump, kbps
LANDSAT	8	256	8	32
SMM	16	512	16	32
ERBS	1.6	32	1.6	-
ST	0.5	1024	4	-
GRO	32	512	1	32
UARS	32	512	1	32
STS	128	128	128	128
TDRS	1	-	1	-

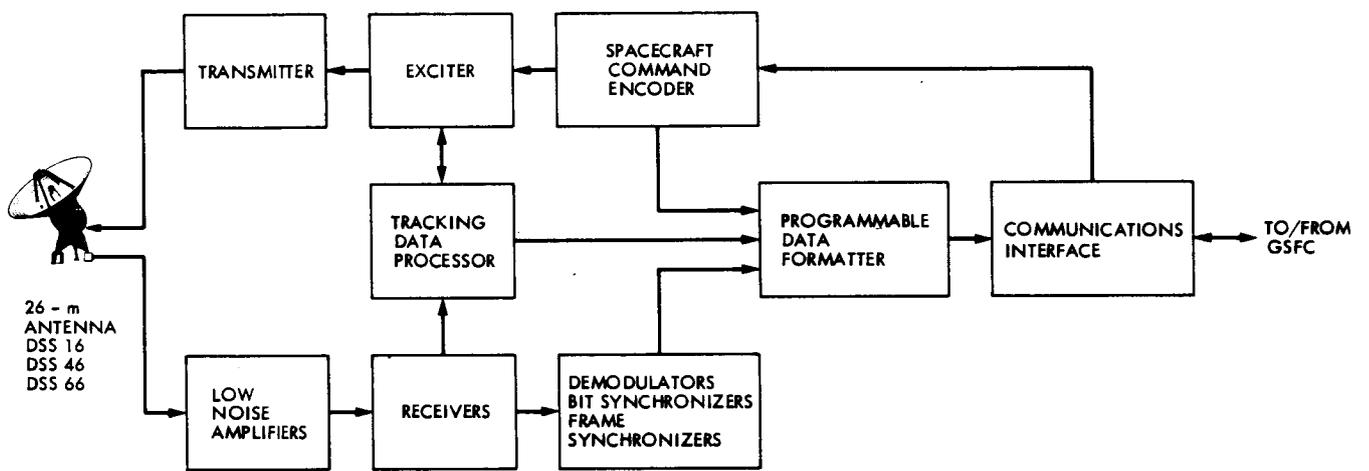


Fig. 1. A typical support configuration

Planning for VLA/DSN Arrayed Support to the Voyager at Neptune

J. W. Layland
TDA Planning Office

D. W. Brown
TDA Engineering Office

Preplanning for the use of the National Radio Astronomy Observatory's Very Large Array (VLA) in support of Voyager at Neptune has been underway since early 1982. When arrayed with the DSN antennas at Goldstone, CA, the VLA more than doubles the potential data return over the American longitude for the 1989 Voyager encounter. This report summarizes the background, rationale, and current status of planning for VLA-DSN Arrayed Support to the Voyager at Neptune.

I. Introduction

The Very Large Array (VLA) is an array of twenty-seven 25-meter antennas in a triradial configuration in the high New Mexico desert. The primary role of this array is developing maps of radio-bright objects in the sky, and it incorporates a large mapping processor which is capable of cross-correlating the 351 ($=27*26/2$) baselines of the array in real time (Ref. 1). One of the optional products of this mapping processor is a combined output which represents the coherent sum of the signals being received at each of the antennas. This combined output of the mapping processor will represent about two and a half of the DSN's 64-meter apertures when all of the VLA antennas are outfitted with X-band low noise amplifiers.

A study of the possible use of the VLA for the support of the Voyager encounter of Neptune in August of 1989 was commissioned by an exchange of letters in early 1982 between Dr. B. C. Murray, then Director of the JPL, and Dr. M. S.

Roberts, then Director of NRAO. The study was handled at JPL concurrent with the somewhat broader Interagency Array Study which considered the total complement of possible non-DSN receiving apertures that might be appropriate for support of the Voyager (Ref. 2). That study reported its conclusions in early 1983, with positive recommendations for the VLA and a few other observatories. Engineering studies have continued since that time, and are reflected in this article.

At the present time, Voyager is virtually unique in the benefits that can be derived by arraying. The reason becomes clear when one realizes that the Voyager was fundamentally designed to give the superb coverage it did at the Jupiter and Saturn encounters, but with a flight path which would much later take it past the far outer planets Uranus and Neptune (Ref. 3). Compensating for the increased distance is being done partially by changes to the on-board control software, but to provide outer planet coverage akin to that at

Saturn will require a substantial increment to ground receiving capability. Equipping a new mission for Voyager-style return at these far outer planets would take many years, and many hundreds of millions of dollars. The flight of the Voyager presents a once-in-a-lifetime opportunity for today's scientists and scientific institutions.

The main message from the study was this: Considering all aspects, including the timeliness and costs of building antenna facilities, and the scientific potential and uniqueness of the Voyager, it is appropriate that most of the additional aperture needed to support the Uranus and Neptune encounters be obtained by enlisting the help of the world's large receiving facilities, providing that such help could be mutually arranged. For the particular case of the VLA, configuring it to provide the equivalent of 2.5 times today's DSN 64-meter antenna capability would boost the receiving aperture over the American longitude so that coverage is available for both the best compressed imaging data rate, and for several of the alternative (higher) data rates without image data compression.

This article updates the VLA material from the Interagency Array Study Report which appeared in *TDA Progress Report 42-74* (Ref. 2).

II. Voyager Mission Characteristics

The objectives of the Voyager mission to Neptune are, generally, to extend the comparative studies of the outer planets to include the environment, atmosphere, surface and body characteristics of the planet Neptune; one or more of its satellites, with emphasis on Triton, and a search for rings. Typical specific scientific objectives to be addressed at Neptune include measurements of the gross morphological structures of the planet and satellites; determination of the Neptune atmospheric composition, structure, and dynamics; determination of the Neptune rotation period; detailed magnetospheric and plasma studies; a study of the satellite surface features, temperatures, and probable Triton atmosphere; and a study of the Neptune rings, if they exist.

The Voyager trajectories, from launch in 1977 onward, are shown in Fig. 1. As can be seen, the decline in Voyager signal strength from Jupiter to Uranus is some 10 dB due to the increase in distance, and the decline from Uranus to Neptune will be another 3.5 dB, most of which must be accommodated on the ground. Figure 2 shows the configuration of the DSN as we expect it to be in 1989, together with the more prominent of the world's apertures which were considered by the interagency array study. It should be noted that the DSN as pictured in this figure is not identical to the network as it will exist in 1986, but includes a per-

formance upgrade of the present 64-meter antennas, and a new high-efficiency 34-meter antenna in Spain.

The telecommunications link performance of the Voyager spacecraft with the 1989 DSN and selected other apertures arrayed is shown in Fig. 3. The horizontal axis of this figure is indexed by GMT hours (Earth-received time) for the day of encounter in August, 1989. The time of closest approach to Neptune is at 08:06, and the time of closest approach to Triton, the largest satellite, is at 13:16 hours. The VLA is well positioned to return the final several hours of Neptune bright-side imaging, as well as to off-load the tape recorder to make room for Triton images. The left axis of this figure is labeled in terms dB of total received signal-to-noise ratio for a reference data rate of 1 bit/s, and the arcs of the figure indicate the effective (normalized) received signal strength. The cruves represent 90% confidence levels, and include the nominal weather statistics and availability of individual antennas. The right-hand axis is labeled to indicate the threshold levels applicable to various Voyager data rates. Communication is feasible at a given rate whenever the received signal strength exceeds the associated threshold level. The information content of the various rates is shown in Table 1. The higher of the two compressed imaging data rates is made feasible by the arraying with two apertures: Parkes Radio Telescope in Australia, and the VLA (2.5 equivalent 64-meter antennas) in the American longitude.

III. Arraying Configuration and Requirements

The functional block diagram for arraying between the DSN Goldstone site and the VLA is shown in Fig. 4. The X-band signal of the Voyager is received concurrently at the DSN facilities and by the observatory. The X-band receiver provides phase-coherent detection of the spacecraft signal, and emits a baseband version of it which contains the same data as the comparable signal at the DSN station. Functional requirements include the provision of a real-time link between the observatory and the DSN site, as well as local recorders at both sites. These would always record the signals, and could enable near-real-time arraying, but would primarily provide backup in case of difficulty with the real-time link. The combiner must be capable of operating with either the link or with the recorders.

With the configuration as shown, the choice is available to operate either using the link as the primary pathway, or with the recorded signal as the primary pathway. Using the real-time link is preferred from an operational standpoint because it provides immediate validation of the array operation, gives the flight project personnel immediate visibility into

their data, and avoids the effort and cost of transporting and processing tapes, unless link outages occur.

Figure 5 shows the currently planned VLA configuration for Voyager support. The items to be added to the VLA include X-band feedhorns, low-noise amplifiers and down-converters at the front ends, and the phase-locked receiver and coherent detector at the combined output of the VLA processor. Use of the VLA's intrinsic signal transmission facility has the advantage that the spacecraft signal reception capability can be tested, and perhaps operated, almost independent of the location of the antennas within the array. It has the disadvantage that the Voyager signal is subjected to a 1.6 ms gap in signal reception per 52 ms control cycle, and to a 3-level quantization in the VLA processor. Analysis has shown the gap to be tolerable when the VLA is arrayed with Goldstone (Refs. 4, 5).

Two channels of the VLA signal transmission and processing equipment may be used: one with a 6 to 12 MHz IF bandwidth to carry the spacecraft data signal at roughly full precision, and a second (not shown in Fig. 5) with narrower bandwidth to be used to self-calibrate the VLA: i.e., to control the differential phase and delay in the system. Careful calibration or control of the differences between these two would be needed. Tests in 1983 had shown that it was feasible to self-calibrate the VLA on a natural signal source at a strength analogous to that of the Voyager spacecraft's signal in the narrow bandwidth (Ref. 6). More recent testing with the Voyager signal itself holds promise that the self-calibration can be achieved with the full bandwidth data channel (Ref. 7).

The baseline design X-band low-noise amplifier is a cryogenically cooled FET amplifier, similar to those in place at the VLA for its other frequency bands. The first two pre-prototype X-band LNAs are now in place on antennas at the VLA, mated with JPL-provided feedhorns, to enable testing of the X-band receive characteristics. A modern alternative to the FET, the so-called HEMT (high electron-mobility transistor), with the potential for a much reduced system temperature, will be evaluated for use in the LNAs (cf. NRAO Voyager Front-End Construction Plan, Rev. A, NRAO internal document, Feb. 15, 1985).

IV. Expected Level of VLA Support

The day of closest approach of the Voyager to the planet Neptune is August 24, 1989 (PDT). At that time, Voyager is visible daily from the VLA for an approximately eight hour "pass." That day, and the several days immediately surrounding it, are the times of greatest importance for the data to be gathered from the encounter. The time intervals for which arrayed support would be requested by the Voyager

project include these critical passes plus several passes for testing the correctness of both configuration and operational procedures plus a number of additional passes both preceding and following the days of close encounter. Times not specifically requested, such as when Voyager is not visible, are presumed to be available for normal VLA astronomy purposes.

It is planned that the VLA will be made available for Voyager-directed support or operational testing for up to 40 observing intervals of up to 8-1/2 hours each. Operational verification tests will begin in April 1989, and continue into May. Actual spacecraft support could begin in mid July, about six weeks prior to encounter, and continue through four weeks after encounter. In addition to the spacecraft operations and operational testing, subsystem testing and verification would be required.

V. Organization Roles

The four organizations involved, NASA, NSF, JPL, and NRAO, will each play a part in preparing the VLA to support the Voyager at Neptune. As the principal beneficiary of this effort, NASA is providing the funding needed to support all of those changes to the VLA which are needed for Voyager. This includes work at both NRAO and JPL. The DSN organization will bear the primary responsibility for system design, planning, and management of the project. Implementation responsibility will be shared between NRAO and JPL. The mechanism implementing these roles is outlined in the "Management Plan for the VLA-GDSCC Telemetry Array Project" (D. W. Brown, JPL Document 1220-1, March 15, 1985 [internal document], Jet Propulsion Laboratory, Pasadena, CA).

With reference to Fig. 5, all equipment which is implemented into the VLA as an integral part of it, such as the X-band LNAs and down-converters, will be instrumented in a way that is suitable for general use by NRAO, and will be retained by NRAO-VLA upon completion of the Voyager support. All equipment which is adjoined to the VLA specifically for the Voyager support, and which is endemic to spacecraft data handling, such as the phase-locked receiver backend, specialized recording capability, satellite link to Goldstone, etc., will be de-implemented by JPL upon completion of Voyager support.

Again with reference to Fig. 5, NRAO will be responsible for the design and implementation of the X-band receiving equipment. This equipment is at least similar to if not identical to comparable elements already a part of the VLA for other frequency bands. The feeds have been designed by JPL for NRAO.

The equipment which is specific for spacecraft support, such as the phase-locked receiver and coherent detector, specialized recording, and communications to Goldstone, will be entirely the responsibility of the DSN, as will the overall operational coordination of the array. The interface between this equipment and the VLA instrumentation will be the single signal combined output of the VLA processor.

VI. Implementation Planning

Instrumentation of the VLA for Voyager support is a long and tedious task because of the large number of antennas involved. Rapid startup is essential if we are not to run afoul of manufacturing, installation, or checkout problems later. The current VLA implementation schedule is shown in Fig. 6. Target date for completion of all installation work is January 1989. The installation of the front-end assemblies throughout the 28 VLA antennas will be allotted two and a half years to provide an orderly process which does not conflict with other necessary VLA activities. Some of the mechanical work for mounting feeds, etc., will take four years because it must take place within the maintenance hangar, and was started at the beginning of 1985. The phase-locked receiver and other back-end elements are planned to be direct extensions of the implementation of the Parkes Radio Telescope for the Voyager encounter with Uranus (Ref. 2).

At a system level, there are tests scheduled in late 1984 and through 1985 which are essential to assuring that the desired capability is achievable. The initial single antenna tests, using a prototype LNA which was developed at the NRAO Charlottesville facility, together with a JPL "spare-parts" feed, have confirmed overall expectations of front end performance as well as provided the first opportunity to view the Voyager spacecraft signal through a part of the VLA's electronics. The dual antenna tests extend the signal path through the VLA correlator/combiner to the output port which would be in use in 1989, albeit with many fewer antennas involved in them. The early lab tests and analysis should support and help the interpretation of the testing at VLA.

A number of review decisions are embedded in the schedule. HEMTs are an improved technology for the FET amplifiers which show promise for significantly reducing their effective temperature, but require evaluation of reliability as well as performance. Back-up power, if deemed necessary, and real-time link equipment would be leased commercial equipment, each requiring appropriate lead time for ordering, contracting, and installation.

VII. Other Users

Other missions and other ground-based radio science users were explored for possible joint interest in a VLA configured so it was capable of 8.4 GHz signal reception. One existing mission, ICE, the International Cometary Explorer, is an S-band mission, and its critical need for large aperture support is occurring in September 1985, long before such capabilities could be available. Of the future mission options examined during the study, most are at modest data rates, and only one, a Titan Probe (1996), showed any benefit from the type of ground aperture increase available through the VLA. Even this benefit seemed minor, and it was concluded that the arrangements with the VLA for Voyager support should be considered as unique to the Voyager, over the next dozen years or so.

There is however additional ground-based science which becomes possible with the X-band capability added to the VLA. The extra frequency, for example, is of some interest in monitoring the spectrum of variable stars and novae, and possibly for Faraday rotation effects in polarization studies. However, there are no spectral lines of more than modest interest within the band, and the planned capability provides no increased sensitivity for detection of continuum sources (A. R. Thompson, NRAO-VLA, private communication, April 23, 1984).

There is some significant interest in the VLA X-band capability for planetary radar, with the DSN Goldstone 64-meter site as transmitter and the VLA as receiver. One instance of this is topographic mapping of the nearer planets such as Venus and Mars. With existing radars such mapping is done by a delay-doppler technique which depends upon some assumptions concerning the general form of the target. The resolution of the VLA would allow these assumptions to be relaxed, thus providing a more direct measurement of the surface shape. Radar observations of more distant targets such as the outer planet satellites and rings would capitalize upon the greater sensitivity of the VLA. Observing proposals for planetary radar observations will be input to the regular VLA proposal evaluation process.

Observations for VLBI at the DSN frequencies for astronomy or other applications could also be of interest. The X-band on the VLA will provide another band in common with the planned Very Long Baseline Array (Ref. 8), which would be useful, though not sufficient by itself to cause it to be implemented.

Acknowledgment

The concept described here for VLA support to the Voyager at Neptune owes a great deal to the support received from the staff at NRAO, and from many members of the JPL Telecommunications Engineering and Voyager Project organizations. Noteworthy here are Dr. A. R. Thompson, Dr. P. Napier, and Dr. S. Weinreb of NRAO, and C. E. Kohlhase, D. A. Bathker, R. C. Clauss, Dr. S. J. Kerridge and B. D. Madsen of JPL.

References

1. Napier, P. J., Thompson, A. R., and Ekers, R. D., "The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope," *IEEE Proceedings*, pp. 1295-1320, Nov. 1983.
2. Layland, J. W., et al., "Interagency Array Study Report," *TDA Progress Report 42-74*, pp. 117-148, Jet Propulsion Laboratory, Pasadena, CA, August 15, 1983.
3. Stone, E. C., "The Voyager Mission: Encounters with Saturn," *Journal of Geophysical Research*, Vol. 88, No. A11, pp. 8639-8642, November, 1983.
4. Deutsch, L. J. "The Performance of VLA as a Telemetry Receiver for Voyager Planetary Encounters," *TDA Progress Report 42-71*, pp. 27-39, Jet Propulsion Laboratory, Pasadena, CA, November 15, 1982.
5. Deutsch, L. J. "An Update on the Use of the VLA for Telemetry Reception," *TDA Progress Report 42-72*, pp. 51-60, Jet Propulsion Laboratory, Pasadena, CA, February 15, 1983.
6. Layland, J. W., Napier, P. J., and Thompson, A. R., "A VLA Experiment - Planning for Voyager at Neptune," *TDA Progress Report, 42-82*, Jet Propulsion Laboratory, Pasadena, CA, this issue.
7. Resch, G. M., *TDA Progress Report*, in preparation.
8. Kellerman, K. I., "The VLBA, Scientific, Technical, and Planning Overview," in *Multidisciplinary Use of the Very Long Baseline Array*, National Academy Press, Washington, D. C., 1983.

Table 1. Voyager Uranus and Neptune data rates

Data Rate, kb/s	Data Type	Equivalent Full Images/hr
29.9	Full frame imaging	13
21.6	Compressed imaging and playback	13 + 6
19.2	Half frame imaging	6
14.4	Compressed imaging	13
11.2	Compressed imaging	9
8.4	Compressed imaging	5
7.2	General science and engineering	None

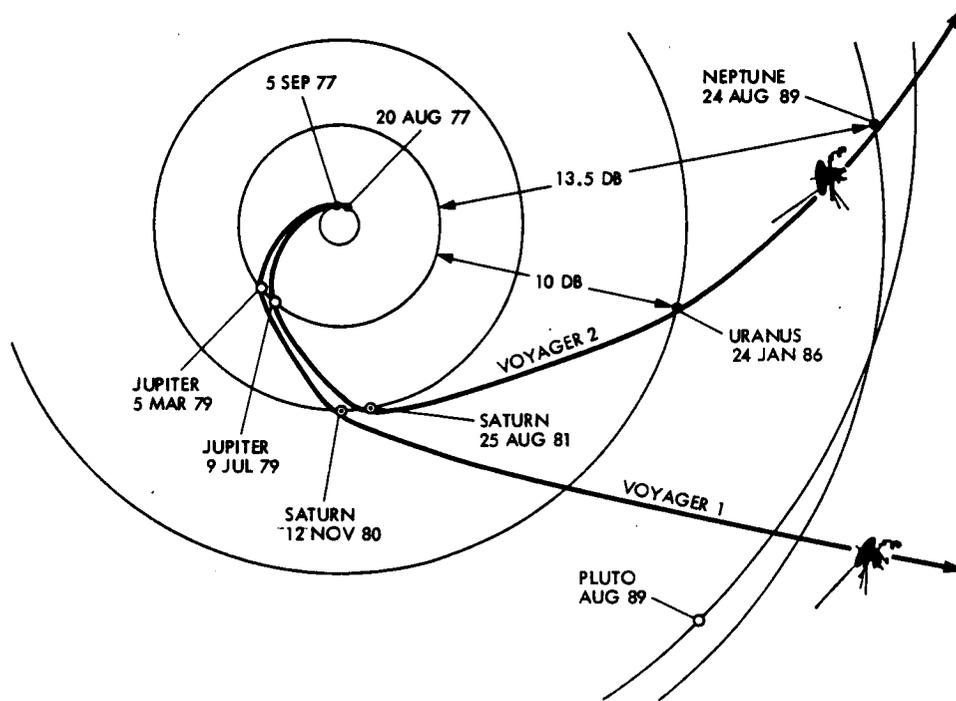
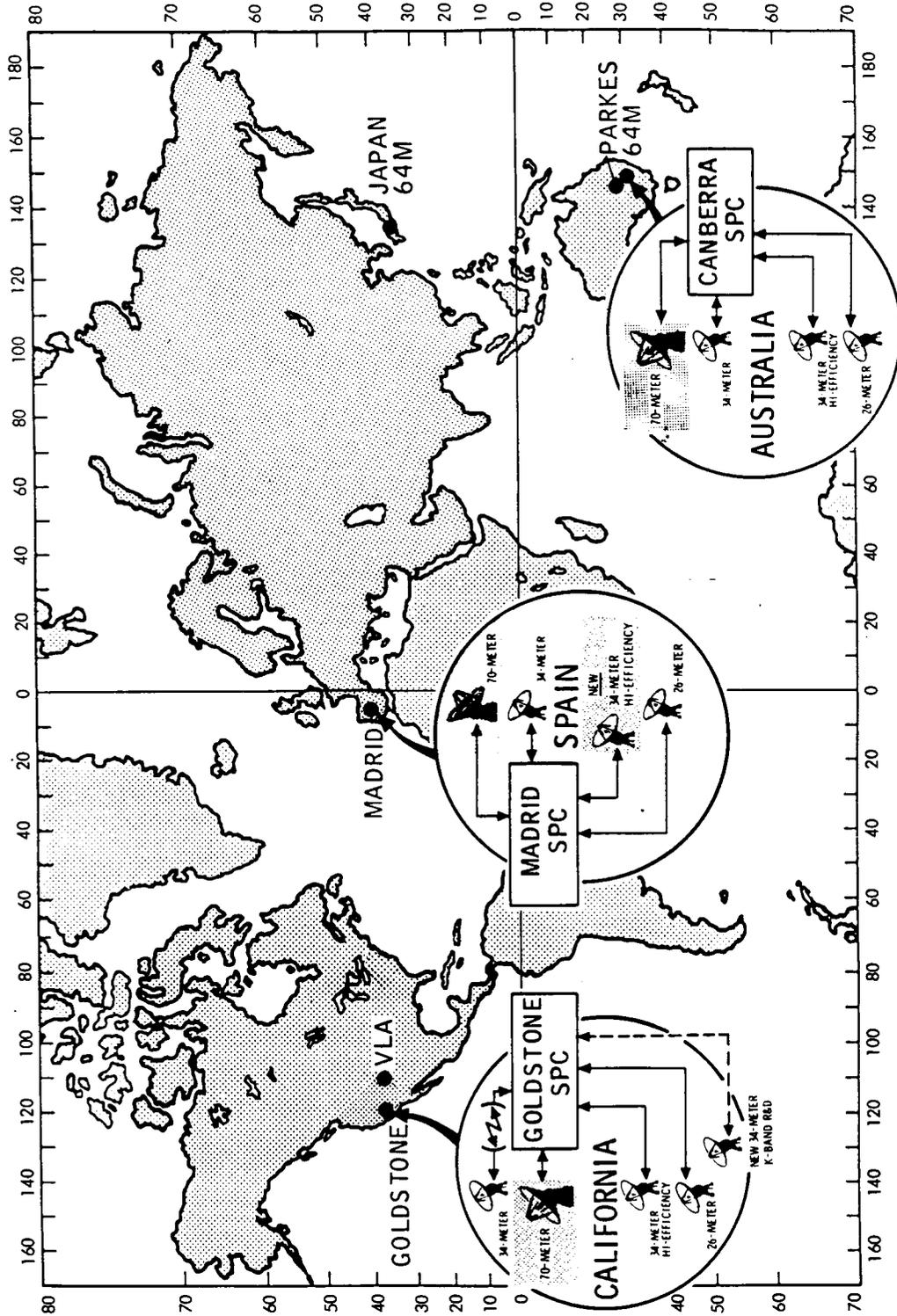


Fig. 1. Voyager trajectory



SHADING INDICATES CHANGES FROM 1986

Fig. 2. 1989 network configuration and candidate facilities

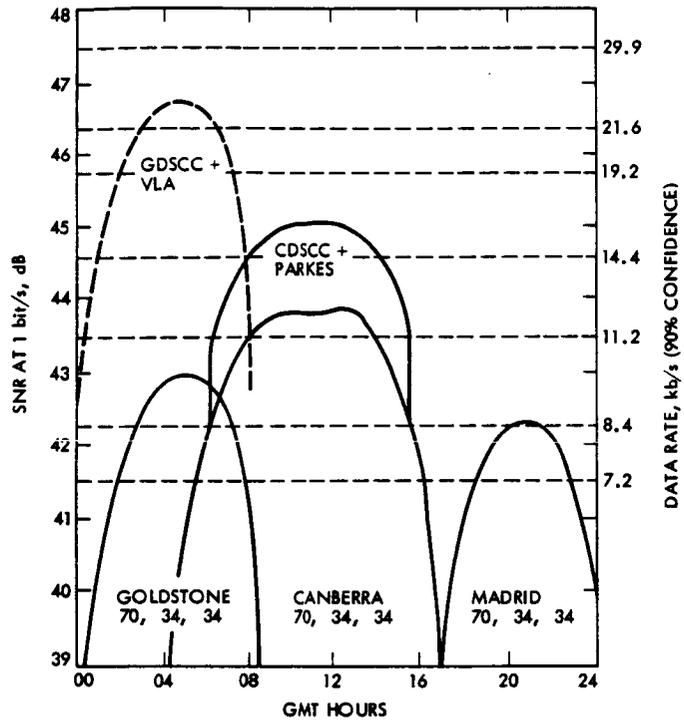


Fig. 3. Telecommunications at Neptune

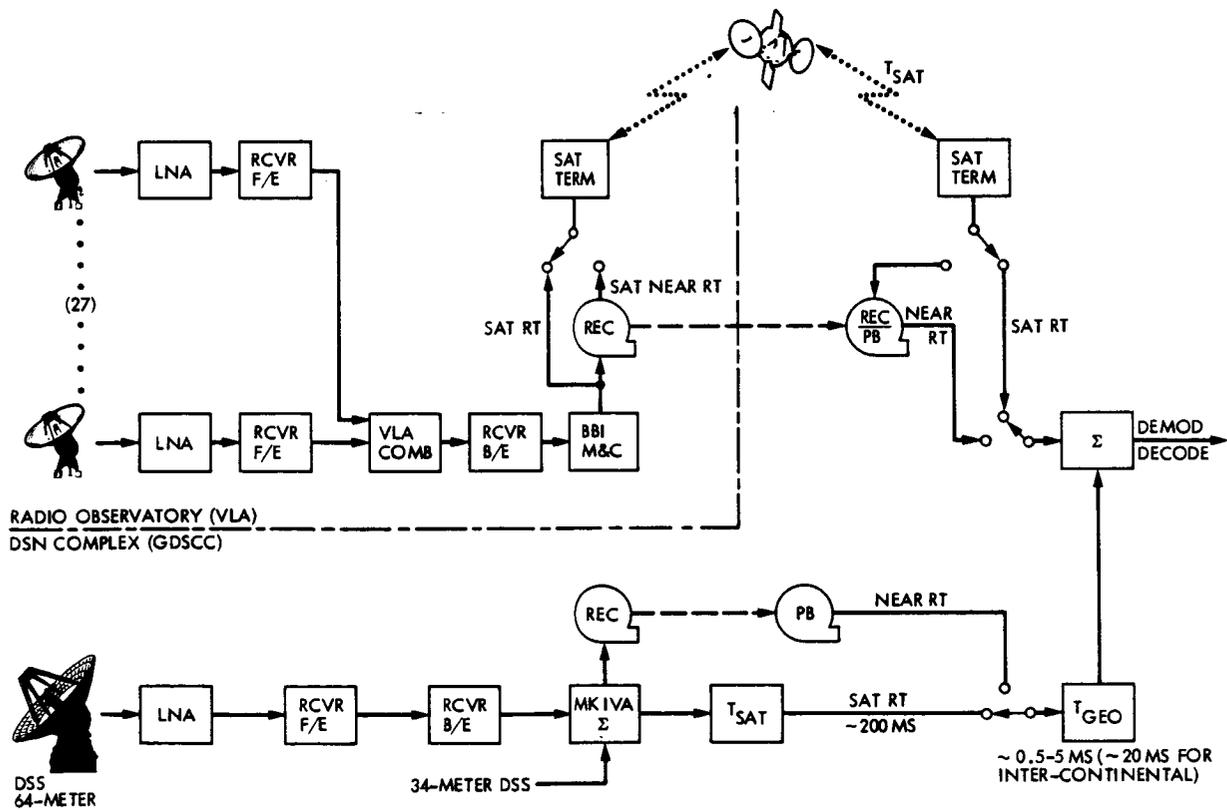
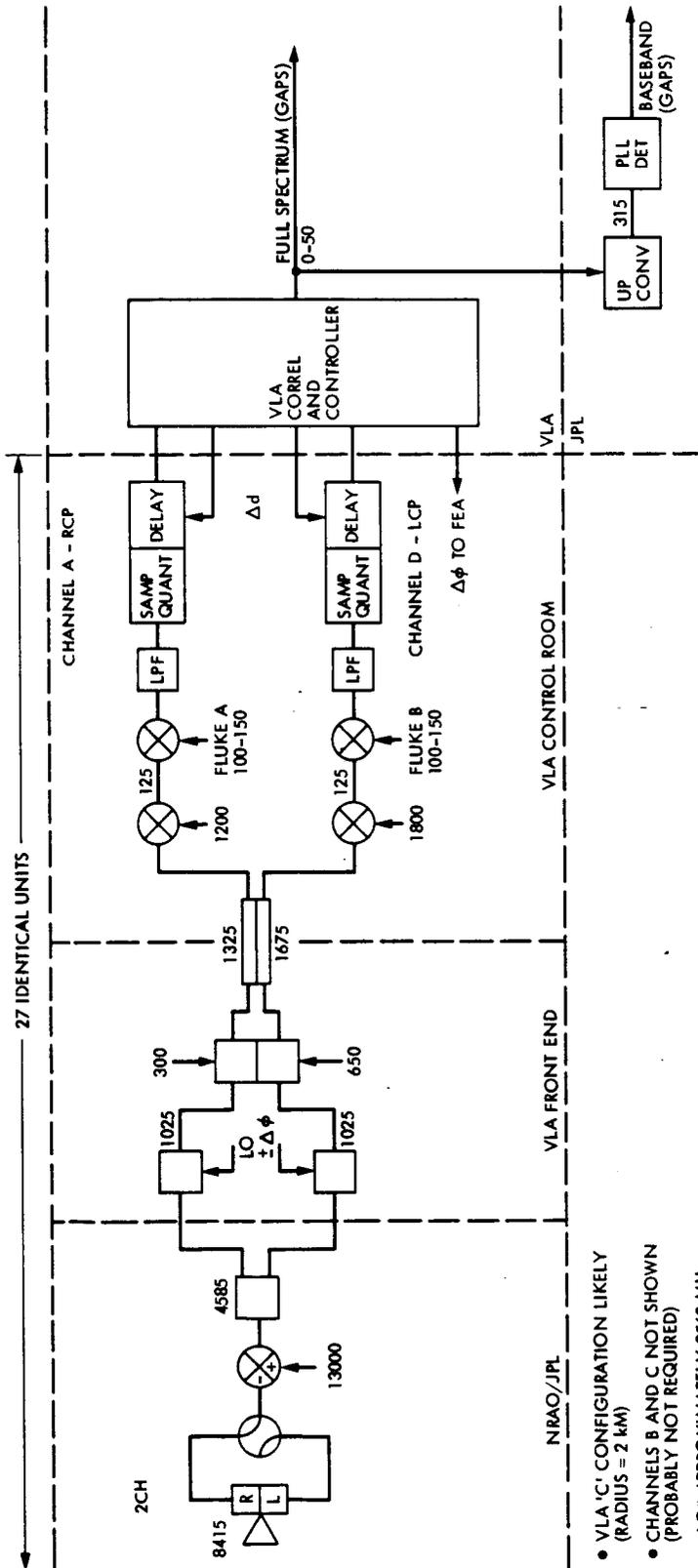
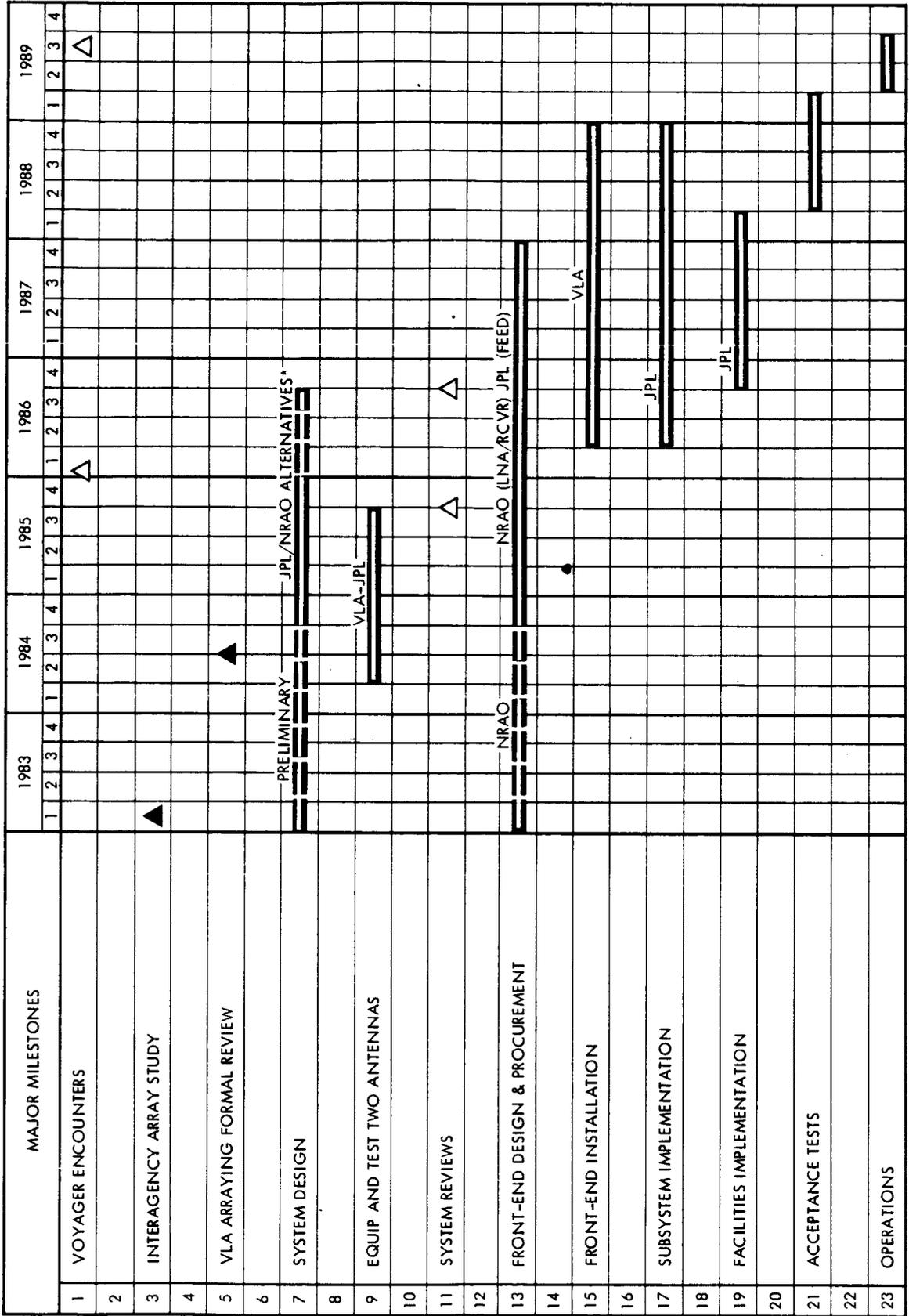


Fig. 4. Arraying functional block diagram



- VLA 'C' CONFIGURATION LIKELY (RADIUS = 2 km)
- CHANNELS B AND C NOT SHOWN (PROBABLY NOT REQUIRED)
- "LO" APPROXIMATELY 3560 MHz

Fig. 5. VLA block diagram



* HEMT VS FET, AUXILIARY POWER, URANUS RECORDING EXPERIENCE

Fig. 6. VLA-GDSCC Telemetry Array project schedule

A VLA Experiment – Planning for Voyager at Neptune

J. W. Layland

TDA Planning

P. J. Napier and A. R. Thompson

National Radio Astronomy Observatory

A VLA engineering experiment was conducted on the night of July 22, 1983 to explore one aspect of the potential for the VLA to support Voyager at its Neptune encounter in August of 1989. Specifically, the experiment tested the ability of the VLA to self-calibrate on a natural radio source whose effective signal strength is the same as Voyager's will be at its Neptune encounter. The experiment was successful and supported the belief that the VLA would be able to be self-calibrated with Voyager's signal.

I. Introduction

The Very Large Array (VLA) is an array of twenty-seven 25-meter antennas in a triradial configuration in the high New Mexico desert. Reference 1 provides a comprehensive overview of the VLA. Details on the VLA, its hardware and software, operation, and the theory behind it may be found in Ref. 2. The primary role of this array is developing maps of radio-bright objects in the sky, and it incorporates a large mapping processor which is capable of cross-correlating the 351 ($= 27 \times 26/2$) baselines of the array in real-time. One of the optional products of this mapping processor is a combined output which represents the coherent sum of the signals being received at each of the antennas. This combined output of the mapping processor would represent about 2-1/2 of the DSN's 64-meter apertures if all of the VLA antennas were outfitted with X-band cryogenic field effect transistors (FET's). As noted in the Interagency Array Study Report (Ref. 3), this capability can be of significant benefit to Voyager at its encounter with Neptune in 1989.

Configuring the VLA to be able to receive signals from Voyager will require significant investments of time and money. One of the many questions asked before seeking to use the VLA was whether the VLA could self-calibrate on the signal from Voyager, and thus be able to easily combine the signals from all 27 antennas without recourse to a stronger nearby natural radio source as a calibrator. As the VLA could not receive at X-band, obtaining a preliminary answer to this question entailed observing some radio source which provided a signal in one of the existing VLA receive bands which was a suitable analogy to the signal expected from Voyager at Neptune in 1989.

Voyager's total signal from Neptune is predicted to be $3.6 \times 10^{-21} \text{ W/m}^2$ (B. D. Madsen, Voyager Telecom, private communication, May 2, 1983). If we filter it into a 60 kHz bandwidth, selecting one sideband and maximizing the band-pass SNR, the signal is cut by 5 dB to $1.1 \times 10^{-21} \text{ W/m}^2$. Since Voyager's signal is matched to the feed polarization and will appear in only one of the two polarizations (right-hand

circular, but not left-hand circular), while the total flux of a natural radio source is measured in both polarizations, the total cataloged flux of the comparable natural radio source is 3.6 Jy. The VLA processor develops integrated cross products each 10 seconds (or a multiple thereof), which are used to estimate the differential phases between antennas. These cross products do not differ in character whether the input bandwidth is 60 kHz or 50 MHz. A natural source of 130 mJy strength in the full VLA bandwidth of 50 MHz should provide the same SNR in the integrated cross-products as a 3.6 Jy source in 60 kHz and, hence, the same as Voyager's signal. This is significantly below the strength at which the VLA is normally calibrated.

In the experiment itself, radio sources at both 0.5 Jy and 0.14 Jy were included in the observation catalog to help expose any sensitivity to source strength at this level. The observing control "deck" was set up using the time sharing PDP-10, and then called by the array operator into the Mod-comp real-time system when needed that evening.

The observation interval occupied 3 hours and contained the observations listed in Table 1. All observations were taken in the 6-cm band, with a nominal system temperature of 60 K. Except where noted, the integration time is 10 seconds and the closed-loop tracking of the differential phases was performed with a loop gain of 1/4. All observations appeared to lock-in solidly, and suggest that self-calibration on still lower source strengths would be quite possible. The estimated performance of the phase tracking operation was extracted by post-processing the recorded data in the VLA computer.

II. Experiment Description

For most of the experiment the VLA was operated with its fundamental integration interval set at ten seconds. Each ten seconds, the 351 complex integrated cross products of the signals received at each of the 27 antennas were recorded for final processing. Also each ten seconds, the 26 cross products for each of the antennas paired with the selected reference antenna were sampled and used to estimate a differential phase which can represent atmospheric shifts, uncompensated equipment drifts, source position offset, etc. When the array is being calibrated, this differential phase is used as the error signal input for a very narrow-band first-order sampled-data phase-locked loop. In normal operation with a 10-sec integration time, the loop gain is set to 1/4, which results in an effective memory time of about 40 sec, or a bandwidth of 0.025 Hz. For the purposes of the engineering experiment being described, the final processing of the recorded data is devoted to estimating how well this sampled-data phase-locked loop was able to track the actual phase variations.

The final non-real-time processing of the data is performed by the VLA program *ANTSOL* (Ref. 2) which uses the stored integrated cross products to perform a global least squares estimate of the apparent gain and phase of each of the 27 antennas. This is the relatively well-known multi-parameter form of the simple least-squares estimator, which will be used but not developed here. The program uses all 351 of the antenna pairs to provide a global solution for all 26 of the phases measured relative to the reference antenna, which is better by a factor of $2/27$ smaller variance than the phase estimates based upon a single integrated cross-correlation. The global solution can be obtained using the elementary (e.g., 10-seconds) integration intervals which were sampled in the real-time processing, or the elementary integrals can be accumulated to increase their SNR before the global solution is done. Assuming there is no significant external perturbation of the relative phases, their sample variance is a measure of the performance of the real-time phase tracking process.

All of the data from the July 23rd engineering experiment were non-real-time processed with an integration time of 120 seconds. A small part of the data was also processed at 30 seconds to improve visibility into the phase-estimator behavior. The approximate relation between the observed phase variance in the final processing, and the variance of the phase attributable to real-time processing is calculated in the Appendix.

III. Observations

As listed in Table 1, the experiment consisted of 10 observations of 6 distinct sources, with distinguishing characteristics to be noted in the following. Sources are identified by their standard cataloged positions, with the first source, 1803+784 being at right ascension of 18 hrs, 03 min, and at declination +78.4 degrees. This source is one of the VLA calibrators, and its recorded flux at 2.52 Jy was used as the reference against which the strengths of other, weaker sources were compared.

The time of this first observation was 04:12 to 04:21 (IAT), and the elevation angle at which it was observed was approximately 45 degrees. Final-processing of this source with a 2-min integration time resulted in an observed phase jitter of 2.5° rms. Converting this into an estimate of the jitter in the real-time control loop by the methods of the Appendix yields a value of 3.7° rms for the phase-tracking process with a 10-sec T_I and loop gain of 1/4. The data from this source was not processed with 30-sec intervals.

For each source, the very beginning of the recorded data is ignored until the acquisition transients have subsided. Then, as many 120-sec (30-sec) segments as will fit into the allotted time are accumulated and the phases and effective gains are

estimated for each of the 27 antennas, using the 351 cross correlations between antenna pairs. The phase parameter of the selected "reference" antenna is arbitrarily fixed at zero. For each segment, the rms of the 26 non-reference phases was calculated. The average, and one-sigma, uncertainty of this rms value was then calculated over the (10 to 20) segments and appears in Table 1 in the columns identified as RMS PHASE. The column identified as SIG-PH is the estimated rms phase variation in the real-time phase tracking process as derived from 120-sec segments by the methods of the preceding section.

The second, third, and fourth observations were all processed with both 120-sec and 30-sec segments. The rms phase at 30-sec integration time is approximately the same as the rms jitter in the real-time control loop. In each case this estimate is larger than that developed on the basis of the 120-sec segments.

These three observations (numbers 2 through 4) were planned to compare the behavior at high (78°) versus low (16°) elevation, and at about 0.5 versus 0.14 Jansky source strengths. It is clear from this set that phase jitter is much more dependent upon elevation angle than upon source strength — at least for the levels at which we need to be concerned. There are presumably variations of the atmospheric path lengths that dominate the observed phase jitter.

Later observations number 7, 8, and 10 used the same 0.14-Jy radio source with varying real-time processing strategies. For number 7, the unit integration time was doubled, to 20 sec, which resulted in an apparent 20% decrease in phase variation, as compared to a 40% decrease which could theoretically have resulted if the source of the phase variation was independent between the 10-sec segments. Further, lengthening the segment time to 60 sec (and correspondingly increasing the loop gain to $1/2$) actually resulted in an apparent 10% increase in phase variations. The expected response if the sources of the phase variations were independent between the 10-sec segments would have been about a 25% decrease.

While these changes are not significant in a statistical sense because of the wide uncertainty of the observed phase variations, they strongly suggest that lengthening the unit observ-

ing time beyond 10 to 20 seconds would filter out some of the real variations. Another way of lengthening the effective measurement for the real-time control operation is to decrease the control loop gain, as was done for observation number 10, where $g = 1/16$. This gain change was effected by changing a Right Arithmetic Shift instruction in the MODCOMP real-time software system. This gain change, by a factor of four, could have effected a factor of two decrease in phase variations if they behaved as independent variables, but only a 20% decrease was observed.

Observations numbers 5 and 6 were "drift" tests, where the VLA was calibrated on the selected radio source at the beginning of the time interval, after which phase control was turned off — the loop gain set to zero — and the phase variations observed as they drifted free over a 15 to 20 minute interval. Starting and ending rms phases listed in Table 1 indicate considerable growth.

Observation number 9 was inserted to test the scaling of source-strength with bandwidth, which was used to set the inferred source strength believed to be Voyager-comparable. The 2.6 Jy source via 3-MHz passband should result in approximately the same phase behavior as 0.5 Jy in 50 MHz. This it did, but the observation was inconclusive because the non-SNR-related effects appear to dominate the phase variations.

IV. Conclusion

Taken together, the several observations in the experiment demonstrate that one should be able to self-calibrate the VLA with a signal comparable in strength to that expected of Voyager at Neptune, without suffering serious SNR loss to the combined signal from the jitter in the real-time phase tracking process. Application of that result to the Voyager signal involved analogies in both the received signal frequency and the VLA processing bandwidths, so further testing of self-calibration with the Voyager spacecraft signal itself will be needed as the X-band capability begins to be assembled on the VLA. The phase variations induced at low elevations are a concern which needs further exploration, as Voyager will be at quite a low elevation throughout its encounter pass.

References

1. Napier, P. J., Thompson, A. R., and Ekers, R. D., "The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope," *IEEE Proceedings*, pp. 1295-1320, Nov. 1983.
2. Hjellming, R. M., Ed., *An Introduction to the NRAO Very Large Array*, 1983 edition, document available from NRAO to users and potential users of the VLA, (National Radio Astronomy Observatory, Socorro, NM 87801).
3. Layland, J. W. et al. "Interagency Array Study Report," *TDA Progress Report 42-74*, pp. 117-148, JPL, August 15, 1983.

Table 1. Summary of observations

OBS No.	Source	Flux	IAT	EL Angle	RMS Phase @ 2 min	Sig-Ph.	RMS Phase @ 30 sec	Notes
1	1803+784	2.52	4:12 - 4:21	45.1 - 45.3	2.5	3.7	-	Calibration
2	1714+219	0.53	4:25 - 4:37	77.8 - 77.5	3.2 ± 0.9	4.8 ± 1.4	6.5 ± 1.0	Ref @ Hi Elevation
3	2037-253	0.57	4:42 - 4:51	15.4 - 17.5	6.0 ± 2.0	9.0 ± 3.0	10.0 ± 2.6	Ref @ Lo Elevation
4	1831-126	0.14	4:57 - 5:09	41.8 - 42.6	3.8 ± 0.9	5.7 ± 1.4	6.3 ± 1.0	Nom. Vgr. Flux
5	1748-253	0.50	5:10 - 5:29	30.5 - 30.1	2 → 5	3 → 7	-	Drift Test No. 1
6	1748-253	0.50	5:30 - 5:47	30.1 - 29.5	4 → 10	6 → 15	-	Drift Test No. 2
7	1831-126	0.14	5:50 - 6:04	43.3 - 43.0	3.4 ± 0.3	4.2 ± 0.36	-	20 sec T_I
8	1831-126	0.14	6:05 - 6:21	43.0 - 42.4	4.2 ± 1.2	4.7 ± 1.3	-	60 sec T_I
9	1908-202	2.6	6:22 - 6:38	35.8 - 35.6	2.7 ± 0.3	4.0 ± 0.45	-	3 MHz Scale
10	1831-126	0.14	6:39 - 6:55	41.4 - 40.2	3.9 ± 0.9	4.1 ± 0.9	-	10 sec @ $g = 1/16$

Appendix

Extraction of Phase Jitter

Let the variance of the elementary phase estimates be denoted as V_{10} , and assume that this is due entirely to thermal noise or other independent noise sources. Assume further that V_{10} is small enough that nonlinearities can be ignored. These estimates are fed in real-time into the first-order phase-locking process with gain $g = 1/4$. The result of this process is that the individual local oscillators of the VLA antennas contain a refined estimate of the differential phase between the reference antenna and each of the others. Let $n_{i,t}$ denote the noise at time t on the phase-difference measured between antenna number i and the reference antenna, and let $\phi_{i,t}$ be the accumulated phase (error) in the antenna number i local oscillator relative to the reference antenna phase. Then assuming the nominal phase to be zero gives

$$\begin{aligned}\phi_{i,t} &= \phi_{i,t-1} + g(n_{i,t-1} - \phi_{i,t-1}) \\ &= g \cdot \sum_{s=-\infty}^{t-1} (1-g)^{t-1-s} \cdot n_{i,s}\end{aligned}$$

for the phase of the i^{th} local oscillator. The variance of the filtered local oscillator phases denoted V_{10f} , is

$$V_{10f} = V_{10} * \sum_{i=0}^{\infty} 9^2 (1-g)^{2i}$$

$$V_{10f} = V_{10} * g/(2-g)$$

$$V_{10f} = V_{10}/7 \text{ for } g = 1/4$$

If the global solution is performed in final processing with the 10-second integrated cross products, the variance of those resultant phase measurements, denoted V_{m10} , is a combination of the residual phase offsets held by the local oscillators, and the additive noise during the measurement interval. In this case the additive noise of the current measurement interval is independent of the actual phase errors which depend only upon past measurement intervals. Let V_{10g} denote the noise variance on the global solution phases based upon a 10-second interval. Then by this independence,

$$V_{m10} = V_{10f} + V_{10g}$$

Although the noises on the 351 cross products are not independent, the individual noise components in the 27 correlated signal streams are. Hence, it is easy to show that the cross products are pair-wise uncorrelated, and therefore the covariance matrix of the inputs to the global least-squares process is diagonal with all non-zero entries equal to V_{10} . The formal covariance matrix of the 26 globally solved phases has all diagonal entries equal to $2=V_{10}/27$ and all non-diagonal entries equal to $V_{10}/27$.

Thus:

$$V_{10g} = \frac{2}{27} * V_{10}$$

Then, solving for the actual real-time phase errors, we find

$$V_{10f} = 0.79 V_{m10}$$

Actual measurement intervals used are 30 sec and 120 sec, in order to reduce the amount of data to be handled to a more manageable level and to reduce the additive noise. As noise samples are independent between each 10-sec sub-interval, adding ' ℓ ' of these together ($\ell = 3$ or 12) will reduce the noise variance by a factor of $1/\ell$ on all cross products (phase-difference measurements) *except* those involving the reference antenna. When phase-difference measurements involving the reference antenna are extended, the filtered local oscillator phase is anticorrelated with respect to the noise in the previous 10-second intervals.

Let $X_{i,t}$ denote the ℓ -interval average of the phase-difference measurement with respect to the reference antenna

$$\begin{aligned}X_{i,t} &= \frac{1}{\ell} \sum_{j=0}^{\ell-1} D(i-r, t-j) \\ &= \frac{1}{\ell} \sum_{j=0}^{\ell-1} \left[n_{i,t-j} - g \sum_{s=-\infty}^{t-j-1} (1-g)^{t-j-1-s} \cdot n_{i,s} \right]\end{aligned}$$

Where $D(i-r, \cdot)$ denotes the 10-second phase-difference measurement, and $n_{i,s}$ as before denotes the additive noise on

that measurement. By collecting like terms, and simplifying, the expression for $X_{i,t}$ becomes

$$X_{i,t} = \frac{1}{\ell} \sum_{j=0}^{\ell-1} n_{i,t-j} (1-g)^j - g \cdot \frac{1}{\ell} \sum_{j=0}^{\ell-1} (1-g)^j \sum_{s=-\infty}^{t-\ell} (1-g)^{t-s} \cdot n_{i,s}$$

The variance of the $X_{i,t}$ is simple to compute as long as the $n_{i,s}$ is independent and becomes

$$V_x = V_{10} \frac{2 - 2(1-g)^\ell}{(2g - g^2) \ell^2}$$

For $\ell = 1$, this expression degenerates to the obvious $V_x = V_{10} + V_{10f}$. For $\ell > 1$, the V_x can be considerably less than V_{10}/ℓ , which is the contribution of the noise alone.

The ℓ -sample average of the local oscillator phases can be extracted from the above expression for the $X_{i,t}$ by removing the additive noise terms from the first summation. Denote the variance of this ℓ -sample averaged phase as $V\{\phi_\ell\}$. This variance is found to be

$$V\{\phi_\ell\} = \frac{V_{10}}{\ell^2} \left[\ell - \frac{2(1-g)[1-(1-g)^\ell]}{(2-g) \cdot g} \right]$$

For $\ell = 1$, this expression degenerates (as it should) to the expression for V_{10f} . Inserting numbers,

$$\begin{aligned} V\{\phi_\ell\} &= 0.060 \cdot V_{10} \quad \text{for } \ell = 12, g = 1/4 \\ &= 0.113 \cdot V_{10} \quad \text{for } \ell = 3, g = 1/4 \end{aligned}$$

Global least-squares estimates of these ϕ_ℓ are extracted, and their statistics computed to derive V_{10f} . Of the 351 baselines used in the global solution, 325 contain additive noise terms which are uncorrelated from each other and from the ϕ_ℓ . The 26 baselines which are used in the real-time phase tracking

process contain noise which is uncorrelated with the noise on the other baselines, but anticorrelated with the associated ϕ_ℓ . Since this apparent suppression of the added noise affects less than 8% of the baselines, and those only partially, we will assume in the following that this correlation between the real-time filtered phase estimates can be neglected, and solve for V_{10f} as determined from the accumulated measurements at 30 and 120 seconds. In retrospect, it is clear that the need for this assumption could have been avoided by discarding the reference antenna and the affected baselines from the global solution process. Using the assumption, we find:

$$V_{10f} = 1.14 V_{m30} \quad \text{using 30 second intervals}$$

$$V_{10f} = 2.27 V_{m120} \quad \text{using 120 second intervals}$$

A few other cases are also of interest, and can be extracted from the foregoing calculations: A) loop gain = 1/16, 10 sec real-time and 120 sec global integrals.:

$$V\{\phi_\ell\} = 0.786 \cdot V_{10f} \quad \text{for 120 seconds, } g = 1/16.$$

Inverting as before,

$$V_{10f} = 1.13 \cdot V_{m120} \quad \text{for } \ell = 12, g = 1/16$$

B) loop gain = 1/4, 20 sec real-time and 120 sec, global integral

$$V\{\phi_\ell\} = 0.619 \cdot V_{20f} \quad \text{for 120 sec, } g = 1/4$$

and

$$V_{20f} = 1.51 \cdot V_{m120} \quad \text{for } \ell = 6, g = 1/4$$

C) loop gain = 1/2, 60 sec real-time and 120 sec global integral

$$V\{\phi_\ell\} = 0.75 \cdot V_{60f} \quad \text{for 120 sec, } g = 1/2$$

$$V_{60f} = 1.24 \cdot V_{m120} \quad \text{for } \ell = 2, g = 1/2$$

These conversion factors will be used to convert from measured phase variations to an estimate of phase variations embedded within the real-time phase-lock process.

Periodic Binary Sequences With Very Good Autocorrelation Properties

S. Tyler and J. Loftsson
Reliability Engineering Section

Computer searches were performed using both an 8086 microprocessor and a Cyber 750 mainframe to find repeated binary phase coded waveforms with very good matched and mismatched autocorrelation properties. The best results for every period up to 64 are given. Sequences with optimal peak sidelobes were discovered for each of these periods. These sequences have extensive applications in radar and communications, particularly in situations when there are very unfavorable signal-to-noise ratios. The best sequence of period 64 when processed using a mismatched filter giving no sidelobes has a reduction in the main lobe of less than 0.23 dB.

I. Introduction

This article discusses binary sequences of lengths up to 64 with very good periodic autocorrelations.

For short periods, optimal sequences are well known. For sequence periods which equal $3 \pmod 4$ and are prime or are of the form $2^n - 1$, direct methods for finding a sequence with, in some respects, an optimal autocorrelation are also known.

However, for most other periods, there is no known practical algorithm for deriving optimal sequences (Ref. 1). Since sequences of lengths of about 40 or more are too long to be subjected to an exhaustive computer search (Ref. 2), smaller searches must be made and practical algorithms developed which may lead to sequences with good periodic autocorrelations.

Periodic sequences such as these have great practical value in radar (Ref. 3) and communications (Ref. 4), especially in situations with extremely adverse signal-to-noise ratios. These sequences also have value in artificial intelligence, since analog solutions and pattern recognition methods may apply and be far superior to digital methods for finding excellent (but not the best) long sequences. In addition, such sequences can be used in cryptography to provide derivable "code books" in situations where "two-key" encryption is not desired. Finally, optimal sequences are ideal for use in searches for extra-terrestrial intelligence; not only are they easy to detect, they also advertise current levels of technology.

One example of the use of such a sequence is the Venus ranging experiment by MIT's Lincoln Laboratory in 1959 and 1961. A binary "pseudorandom" shift register sequence of period $2^{13} - 1 = 8191$ was used to determine whether to

transmit "pulse" or "no pulse" in consecutive time intervals. This sequence was very easy to synthesize and had the property that its autocorrelation was recoverable despite a noise-to-signal excess of many decibels (Ref. 5). Due to the difficulty of analyzing an 8191-sequence, planetary ranging sequence periods are generally shorter ($2^8 - 1 = 255$ is common) or at least factorable. Even a sequence of period 8192 would be relatively easy to analyze using Fast Fourier Transform techniques. However, the computers of 1959 did not have sufficient capability to permit the synthesis of an adequate 8192-sequence. The ranging systems used at the Jet Propulsion Laboratory have tended to use Boolean combinations of several shorter sequences to facilitate rapid acquisition. The combination is used to specify phase modulation on a continuous wave carrier; this technique requires less maximum power output than does amplitude modulation (Ref. 5).

This article will discuss the following:

Matched Periodic Binary Sequences: A description and example of how to evaluate the autocorrelation of a sequence.

Calculation of Mismatched Values: A description and example of how to calculate the main-lobe loss when a sequence is analyzed by a mismatched filter.

Sequence Generation Techniques: A description of the techniques used to generate good sequences.

Results: Tables of the best values found, both for matched values and those analyzed by a mismatched filter.

Some of the values listed are "optimal"; others are merely the best the authors have been able to obtain to date. The main emphasis has been on finding the best value for a sequence of period 64 — the best previous value was improved by over 27%.

II. Matched Periodic Binary Sequences

A binary sequence (or binary code) is a string of bits. It can be thought of as a vector, c , where each c_i is a plus one or a minus one. A periodic sequence is one which is continuously repeated; for a binary sequence of period j , $c_{\ell+j} = c_\ell$ for all ℓ .

The "autocorrelation," a , of a sequence c of period j is:

$$a_\ell = \sum_{k=1}^j c_k c_{k+\ell}$$

When the autocorrelation is normalized by dividing it by j , it is called the "autocorrelation function."

Here a_j, a_{2j}, a_{3j} , and so forth are "main lobes," the remaining a_ℓ are "sidelobes;" a is considered to have j "elements," one main lobe and $j - 1$ sidelobes.

For a sequence to have "good matched autocorrelation properties," it must satisfy at least one of the following criteria:

- (1) The peak sidelobe in the autocorrelation is small.
- (2) The sum of the squares of the sidelobes in the autocorrelation is small.

These concepts are illustrated by means of an example. Consider a sequence of period 7:

- - + + - + +

To get the elements of the autocorrelation, suppose the following:

- - + + - + + is the original sequence. Then
+ - - + + - + is the sequence shifted one position.

- + - + - - + is the arithmetic product for each position. The number -1 is the sum of these products; it is the first sidelobe element of the autocorrelation.

Shifting by 2,

- - + + - + +
+ + - - + + -

- - - - - + -

The number -5 is the sum; it is the second sidelobe element of the autocorrelation.

Shifting by 3,

- - + + - + +
- + + - - + +

+ - + - + + +

The number +3 is the sum and the next element in the autocorrelation.

Shifts by 4, 5, and 6 positions are equivalent to those of 3, 2, and 1. The last element of the autocorrelation is the main lobe. It corresponds to the original unshifted sequence. The other elements are the sidelobes (the main lobe is not a sidelobe). Thus, the autocorrelation of - - + + - + + is -1, -5,

3, 3, -5, -1, 7. Here $P = \text{peak sidelobe magnitude} = 5$; $M = \text{sum of squares of sidelobes} = 70$.

An "optimal" sequence for period 7 is $+++--+-$ and has the autocorrelation $-1, -1, -1, -1, -1, -1, 7$, where $P = 1$, $M = 6$.

For period 8, $++++-+-$ is an optimal sequence. It has the autocorrelation $0, 0, 0, -4, 0, 0, 8$. Here $P = 4$, $M = 16$.

The remainder of this section shows the connection between the autocorrelation and the Fourier transform.

As a preliminary, it should be noted that $a_\ell = a_{j+\ell} = a_{j-\ell}$ for all ℓ :

$$\begin{aligned} a_{j+\ell} &= \sum_{k=1}^j c_k c_{k+j+\ell} = \sum_{k=1}^j c_k c_{k+\ell} = a_\ell = \sum_{k=\ell+1}^{j+\ell} c_{k-\ell} c_k \\ &= \sum_{k=1}^j c_{k-\ell} c_k = \sum_{k=1}^j c_{k+j-\ell} c_k = a_{j-\ell} \end{aligned}$$

It is useful to have a matrix \mathbf{Z} which satisfies $\mathbf{a} = \mathbf{Z}\mathbf{c}$, that is:

$$a_\ell = \sum_{k=1}^j Z_{k\ell} c_k$$

However, the "circular convolution matrix" \mathbf{R} of the sequence \mathbf{c} is actually a $j \times j$ matrix satisfying:

$$a_{\ell-1} = \sum_{k=1}^j R_{k\ell} c_k$$

Since

$$\begin{aligned} a_{\ell-1} &= a_{\ell+j-1} \\ &= \sum_{k=1}^j c_k c_{k+\ell+j-1} \\ &= \sum_{k=1}^j c_k c_{k+1-\ell} \end{aligned}$$

Then $R_{k\ell} = c_{k+1-\ell}$

The "Fourier transform" λ of the sequence \mathbf{c} , a vector $\lambda = \mathbf{D}\mathbf{c}$, satisfies:

$$\lambda_\ell = \sum_{k=1}^j D_{k\ell} c_k$$

where

$$D_{k\ell} = \frac{1}{\sqrt{j}} \omega^{(k-1)(\ell-1)}$$

$$\omega = \exp(2\pi i/j)$$

$$i = \sqrt{-1}$$

The sequence \mathbf{c} can be restored from λ by means of the "inverse Fourier transform":

$$c_k = \sum_{m=1}^j D_{km}^* \lambda_m^*$$

where

λ_m^* = the complex conjugate of λ_m

$$D_{km}^* = \frac{1}{\sqrt{j}} \omega^{j-(k-1)(m-1)}$$

The circular convolution matrix can therefore be expanded in terms of an inverse Fourier transform:

$$\begin{aligned} R_{k\ell} = c_{k+1-\ell} &= \frac{1}{\sqrt{j}} \sum_{m=1}^j \omega^{j-(k+1-\ell-1)(m-1)} \lambda_m^* \\ &= \frac{1}{\sqrt{j}} \sum_{m=1}^j \omega^{j-(k-1)(m-1)} \lambda_m^* \omega^{(m-1)(\ell-1)} \\ &= \sqrt{j} \sum_{m=1}^j D_{km}^* \lambda_m^* D_{m\ell} \end{aligned}$$

Thus

$$\mathbf{R} = \sqrt{j} \mathbf{D}^* \boldsymbol{\Lambda}^* \mathbf{D}$$

where

$$\Lambda_{ij}^* = \delta_{ij} \lambda_i^*$$

$$\begin{aligned} \delta_{ij} &= 1 & i=j \\ &= 0 & i \neq j \end{aligned}$$

The criterion of optimality that average sidelobe response be minimized with respect to mainlobe response means minimizing

$$\frac{\sum_{\ell=1}^{j-1} a_{\ell}^2}{a_j^2}$$

Since a_j^2 always equals j^2 , this is equivalent to minimizing the sum of the squares of the sidelobes:

$$\sum_{\ell=1}^{j-1} a_{\ell}^2$$

which is in turn the same as minimizing

$$\begin{aligned} \sum_{\ell=1}^j a_{\ell}^2 &= \sum_{\ell=1}^j a_{1-\ell}^* a_{1-\ell} = c^* R^* R c \\ &= j c^* D^* |\Lambda|^2 D c = j \lambda^* |\Lambda|^2 \lambda \end{aligned}$$

This is equivalent to minimizing

$$\sum_{i=1}^j |\lambda_i|^4$$

If evaluations of sequences are to be performed on a computer in a language which includes no bit manipulation instructions, calculating the λ_i is more efficient than evaluating the autocorrelation. It requires j^2 multiplications to calculate the autocorrelation unless individual bits are used. To calculate the λ_i using a Fast Fourier Transform (FFT) requires fewer than $j (\log_2 j)$ operations. However, in an assembly language, a maximum of $j + (j/\text{word length})$ logical operations are needed to replace the j^2 multiplications in calculating the autocorrelation.

When calculating the λ_i , it is helpful to check the normalization. By Parseval's theorem:

$$\sum_{i=1}^j |\lambda_i|^2 = j$$

III. Calculation of Mismatched Values

When sequences are used for ranging, they are phase coded rather than amplitude modulated. Since the signal-to-noise ratio (SNR) is expected to be very low, it is generally favorable to use maximum amplitude throughout; amplitude modulation would be inconsistent with this requirement. Similarly, periodic rather than aperiodic waveforms are generally used to increase the redundancy of the information. A single (aperiodic) sequence uses 2^n bits to transmit only n bits of information (the information transmitted being the displacement of the starting point of the sequence). However, a repeating (periodic) sequence uses $m \times 2^n$ bits to transmit the same information, where m is the number of repetitions which are processed (whether this is truly the best way to use $m \times 2^n$ bits to transmit n bits of information in a noisy environment is not the issue).

The detection procedure is equivalent to comparing the incoming signal to a template consisting of the original sequence and "moving the template around" until it matches the signal. In this situation the autocorrelation is the "output" of a "matched receiver."

Although the signal is not amplitude modulated, the *template* may have some amplitude modulation. The incoming sequence is then no longer correlated with itself but with a similar "weighted" sequence. By varying the amplitude of each bit in this template sequence, the sidelobes can be reduced or even eliminated.

The cross-correlation, x , of two sequences, b and c , each of period j is:

$$x_{\ell} = \sum_{k=1}^j b_k c_{k+\ell}$$

Let c be a binary sequence which is to be cross-correlated with b , a sequence composed of real numbers. Sequences c and b are related by the "weighting function" t .

$$b_{\ell} = t_{\ell} c_{\ell}$$

In this case, b can be considered a "mismatched filter" to c . To normalize this mismatched filter:

$$\sum_{\ell=1}^j b_{\ell}^2 = j$$

For a given sequence, there usually exists a mismatched filter which can be used to mathematically operate on the sequence so as to reduce all the sidelobe values to zero. For $j > 4$, however, the main-lobe value will also be reduced somewhat. The sequence with the best mismatched (cross-correlation) properties is the one which has the smallest decrease in main-lobe value and therefore has the smallest ratio of

$$\frac{j^2}{x_j} = \frac{j^2}{\sum_{k=1}^j b_k c_k}$$

The circular convolution matrix R for mismatched sequences is:

$$x_{\ell-1} = \sum_{k=1}^j R_{k\ell} b_k$$

The average-to-peak cross-correlation response is then:

$$\frac{\sum_{i=1}^{j-1} x_i^2}{x_j^2}$$

Minimizing the above is equivalent to minimizing

$$\frac{b^* R R^* b}{b^* c c^* b}$$

which occurs when

$$b = (R R^*)^{-1} c$$

$$= \frac{1}{j} D^* |\Lambda|^{-2} D c$$

This gives b_k proportional to

$$\sum_{\ell=1}^j D_{k\ell}^* \frac{1}{\lambda_\ell^*}$$

This choice of b zeros the cross-correlation (Ref. 5). The only constraint necessary for $R R^*$ to be nonsingular is that $|\lambda_i|^2 > 0$ for all i , i.e., that $R R^*$ is positive definite.

The best sequence is the one which minimizes

$$\frac{1}{j} \sum_{i=1}^j \frac{1}{|\lambda_i|^2}$$

In practice, smaller weights may be chosen to reduce the SNR loss. In this case, the sidelobes will be reduced rather than eliminated.

The following example illustrates the derivation of the appropriate mismatched filter for a given sequence.

Consider the sequence of length 8 discussed in the previous section. The sequence $+++ + - - -$ has autocorrelation $0, 0, 0, -4, 0, 0, 0, 8$, where $P =$ peak sidelobe magnitude = 4 and where $M =$ sum of squares of sidelobes = 16.

It will be shown that when this sequence is analyzed with the appropriate mismatched filter, b , the cross-correlation becomes $(0, 0, 0, 0, 0, 0, 0, x_j)$.

The elements of the mismatched filter, b_k , must be normalized so that

$$\frac{1}{8} \sum_{k=1}^8 b_k^2 = 1$$

in order to obtain the correct value of x_j .

The loss in SNR for the mismatched filter is then $L = (8/x_j)^2$. To actually calculate L :

$$L = \frac{1}{8} \sum_{k=1}^8 \frac{1}{|\lambda_k|^2}$$

where the λ_k are elements of the Fourier transform of the original sequence:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} = \sqrt{\frac{1}{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega & i & -\omega^* & -1 & -\omega & -i & \omega^* \\ 1 & i & -1 & -i & 1 & i & -1 & -i \\ 1 & -\omega^* & -i & \omega & -1 & \omega^* & i & -\omega \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -\omega & i & \omega^* & -1 & \omega & -i & -\omega^* \\ 1 & -i & -1 & i & 1 & -i & -1 & i \\ 1 & \omega^* & -i & -\omega & -1 & -\omega^* & i & \omega \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

where

$$i = \sqrt{-1} = \exp(i\pi/2)$$

$$\omega = \sqrt[4]{-1} = \exp(i\pi/4) = (1+i)/\sqrt{2}$$

$$\lambda = \frac{1}{\sqrt{8}} \begin{bmatrix} 2 \\ 2(1+i) - \sqrt{2}(1-i) \\ 2i \\ 2(1-i) + \sqrt{2}(1+i) \\ -2 \\ 2(1+i) + \sqrt{2}(1-i) \\ -2i \\ 2(1-i) - \sqrt{2}(1+i) \end{bmatrix}$$

Thus

$$|\lambda|^2 = \frac{1}{8} (4, 12, 4, 12, 4, 12, 4, 12)$$

$$\frac{1}{|\lambda|^2} = (2, 2/3, 2, 2/3, 2, 2/3, 2, 2/3)$$

$$\begin{aligned} L &= \frac{1}{8} \sum_{k=1}^8 \frac{1}{|\lambda_k|^2} \\ &= \frac{1}{8} (8 + 8/3) \\ &= 4/3 \\ &= 1.3333 \end{aligned}$$

The loss in dB is $10 \log_{10} (1.3333) = 1.250$ dB.

For this particular case, the mismatched filter elements can be found by inspection. They also can be calculated as follows:

$$b_i \propto \sum_{k=1}^j D_{ik}^* \frac{\lambda_k}{|\lambda_k|^2}$$

where D_{ik}^* is the inverse Fourier transform.

There is no need to normalize the b_i at this point; it can always be done later since

$$\sum_{i=1}^j b_i^2 = j$$

The filter \mathbf{b} is proportional to

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega^* & -i & -\omega & -1 & -\omega^* & i & \omega \\ 1 & -i & -1 & i & 1 & -i & -1 & i \\ 1 & -\omega & i & \omega^* & -1 & \omega & -i & -\omega^* \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -\omega^* & -i & \omega & -1 & \omega^* & i & -\omega \\ 1 & i & -1 & -i & 1 & i & -1 & -i \\ 1 & \omega & i & -\omega^* & -1 & -\omega & -i & \omega^* \end{bmatrix} \begin{bmatrix} 6 \\ 2(1+i) - \sqrt{2}(1-i) \\ 6i \\ 2(1-i) + \sqrt{2}(1+i) \\ -6 \\ 2(1+i) + \sqrt{2}(1-i) \\ -6i \\ 2(1-i) - \sqrt{2}(1+i) \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 3 \\ 1 \\ 1 \\ -1 \\ 3 \\ -1 \\ -1 \end{bmatrix}$$

With \mathbf{b} properly normalized:

$$\mathbf{b} = \frac{1}{\sqrt{3}} (1, 3, 1, 1, -1, 3, -1, -1)$$

It is easily verified that the cross-correlation of

$$(1, 1, 1, 1, -1, 1, -1, -1)$$

and

$$\frac{1}{\sqrt{3}} (1, 3, 1, 1, -1, 3, -1, -1)$$

is

$$(0, 0, 0, 0, 0, 0, 0, 4\sqrt{3})$$

IV. Sequence Generation Techniques

A. Iterative Improvement

The basic concept in deriving a good sequence is to start with an arbitrary sequence of period j and perturb one bit of the sequence to create a new sequence. Then the criteria of optimality developed in the previous section are used to decide whether the new sequence is superior to the old one. If the new sequence is inferior, a new bit of the old sequence is perturbed; if the new sequence is superior, it replaces the old sequence and is itself then modified by one bit. Eventually a "locally optimal" sequence is thus obtained. A new initial sequence is chosen, and the procedure is repeated as often as desired.

Perturbing a sequence by 2 bits or 3 bits was also tried; it yielded significantly inferior results to 1-bit modifications.

It may appear that one should look at, say, all ${}_jC_4$ sequences which vary by four bits from the best sequence

obtained by iterative improvement once one-bit, two-bit, and three-bit modifications have failed to improve it rather than abandon the sequence and start over. Certainly, it can not hurt to apply this technique to the best sequence ever found. However, in general, the time spent in an often futile attempt (via 4-bit modifications) to improve one already "locally optimized" sequence could be better spent by locally optimizing a thousand new sequences from scratch.

The iterative procedure also produces good sequences faster than does an exhaustive search. For $j = 43$, at least one out of every 10^7 sequences examined had $M = 42$; for an exhaustive search, only one out of every 10^{10} sequences would have that value. Even if the exhaustive search examined sequences only when

$$\sum_{i=1}^j c_i = \pm 1$$

only one of 10^9 sequences would have $M = 42$. An actual exhaustive search might be restricted to sequences beginning with two or more +1's followed by a -1; this would not increase the rate of finding $M = 42$ sequences.

The best results were obtained by minimizing M , but good results should also be obtainable by minimizing L . When iteratively improving a sequence by calculating L , one need not continually recalculate the sum in

$$\lambda_\ell = \sum_{k=1}^j D_{k\ell} c_k$$

If element c_n of the original sequence is to be perturbed, the new Fourier transform elements, λ'_ℓ , are always simply:

$$\lambda'_\ell = \lambda_\ell \pm 2D_{n\ell} = \lambda_\ell \pm (2/\sqrt{j}) \exp(2\pi i(n-1)(\ell-i)/j)$$

which is noticeably more efficient, especially when j is prime so that an FFT does not help.

It is also possible to modify a sequence iteratively by noting which elements of the Fourier transform are farthest from unity and then perturbing the appropriate bit or bits in the sequence to improve these worst values of λ_ℓ .

B. Choosing an Initial Sequence

Since numerous sequences were to be chosen, an important criterion was to avoid accidentally repeating initial sequences. This was done by making each 4-bit "nibble" of the first initial sequence different and then systematically changing the sequence of nibbles. Rotations of initial sequences are

unlikely to give identical results after iteration (but ones complementation will). Thus for $j = 64$, this procedure can supply $16!/2 > 10^{13}$ initial sequences while for $j = 48$ it gives $16!/(2 \times 4!) > 4 \times 10^{11}$ initial sequences.

Attempts to improve on the choice of initial sequences by modifying sequences of periods $j \pm 1, j \pm 4, 2j, j/2, j/4, \sqrt{j}, j_1 + j_2 = j$, and so forth (where the unmodified sequence had good autocorrelation properties) did not yield better results. It is also difficult to synthesize a large number of such initial sequences.

It may seem that a minor modification of a sequence with, say, $j = 1023$ and $M = 1022$ will give a good sequence with $j = 1024$; this is simply not true. An even worse idea would be to create a $j = 1023$ sequence with a shift register and then pretend it has $j = 1024$ and analyze it with a Fast Fourier transform.

C. Methods of Finding Good Sequences by Inspection

1. **Quadratic residue sequences.** For odd prime periods, j , one forms a quadratic residue sequence by setting to -1 all elements c_i for which $i = n^2 \pmod j$ for some integer $n < j/2$. The remaining elements are set to +1. These sequences give $M = j - 1$ (always optimal) for $j = 3 \pmod 4$. For $j = 1 \pmod 4$, half of the a_i equal +1 and the remaining a_i are -3, thus $M = 5(j - 1)$ which presumably is never optimal, merely good.

For example let $j = 11$. Elements 1, 4, 9, 5 = 16 mod 11, and 3 = 25 mod 11 are set to -1. The remaining elements are +1. This gives a sequence with $M = 10$.

An integer I is a quadratic residue modulo n if $m^2 = I$ (modulo n) has a solution for some integer m and $(I, n) = 1$. When p is an odd prime, the Legendre symbol (I/p) is defined as:

$$\left(\frac{I}{p}\right) = \begin{cases} 1 & \text{if } I \text{ is a quadratic residue modulo } p \\ -1 & \text{otherwise} \end{cases}$$

For this reason, quadratic residue sequences are also referred to as Legendre sequences.

Quadratic residue sequences always have sums

$$\sum_{i=1}^j c_i = \pm 1$$

As will be shown later, such sequences do not have good mismatched autocorrelation properties. It would be more useful to discover an algorithm (if one exists) to produce the

An $M = 192$ sequence for $j = 52$ can be obtained by multiplying the 4 and 13 sequences. The 4-sequence, \mathbf{d} , is: $d_1 = 1, d_2 = 1, d_3 = 1, d_4 = -1$, and $d_i = d_{i-4}$. In hex notation, replacing the -1 by a 0, this is E. Repeating this thirteen times, $\mathbf{d} = \text{EEEEEEEEEEEEEE}$. The 13-sequence, \mathbf{y} , is, in hex notation: 1F35. Repeating this four times, $\mathbf{y} = \text{F9AFCD7E6BF35}$. Then $z_i = d_i y_i$ gives $\mathbf{z} = \text{E8BEDC6F7AE24}$ which has $M = 192$ as well as an L equal to that of the 13-sequence, namely 1.040. The "exclusive nor" (the ones complement of the "exclusive or") operation is used to perform this multiplication in hex notation.

5. Golay sequences. There is another method to obtain an $M = 192$ sequence for $j = 52$.

Let $\mathbf{y} = \text{F9AFCD7E6BF35}$, as before. Then let

$$\begin{aligned} d_i &= 1 & 1 < i < 26 \\ d_i &= -1 & 27 < i < 39 \\ d_i &= 1 & 40 < i < 52 \end{aligned}$$

In hex, $\mathbf{d} = \text{FFFFFFFFC001FFF}$

$$z_i = d_i y_i \text{ now gives } \mathbf{z} = \text{F9AFCD4195F35}$$

which is a sequence of the form $\mathbf{S S S S}$ where each \mathbf{S} is a sequence of length 13. The sequence \mathbf{z} has $L = 1.050$. Sequences of the more general form $\mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_1 \mathbf{S}_2$ are called Golay sequences. Sequences of the form $\mathbf{S S S S}$ do not generally have the property of producing good autocorrelations. However, they do have the property that, when \mathbf{S} is of length ℓ , the first $\ell = j/4$ sidelobes on either side of the main lobe are 0.

D. Demonstration of Equivalence

Two sequences which are *equivalent* always have the same M value (the converse is not true). Two sequences, \mathbf{c} and \mathbf{d} , can be shown to be equivalent, $\mathbf{c} \sim \mathbf{d}$, as follows:

- (1) $\mathbf{c} \sim \mathbf{c}$.
- (2) If $\mathbf{c} \sim \mathbf{e}$ and $\mathbf{e} \sim \mathbf{d}$, then $\mathbf{c} \sim \mathbf{d}$.
- (3) If $\mathbf{c} \sim \mathbf{d}$, $\mathbf{d} \sim \mathbf{c}$.
- (4) If $c_i = -d_i$ for all i then $\mathbf{c} \sim \mathbf{d}$.
- (5) If $c_i = d_{i+k}$ for all i and some k , then $\mathbf{c} \sim \mathbf{d}$.
- (6) If $c_i = d_{i \times k}$ for all i and some k where $(i, k) = 1$, then $\mathbf{c} \sim \mathbf{d}$.

As a special case, when $k = j - 1$, $c_i = d_{j-i}$, a "mirror image" sequence.

When an exhaustive search is made, only one sequence from each equivalence class need be examined: an efficient algorithm for performing such a search is not known.

One method for restricting a search to a small number of members of each equivalence class is to examine only sequences beginning with several +1's followed by a -1; it is not evident that any significant advantage can be obtained in this manner, however.

E. Proof of Optimality

All of the best values of P discovered for each j as well as a number of M values are given as optimal. Most of the proofs of optimality are trivial and based only on the fact that when $j = N \bmod 4$, each element of the autocorrelation is also $N \bmod 4$. For example, when $M = j - 1$ for odd j , one may be sure that no smaller value of M can be obtained. Similarly, $M = 4(j - 1)$ must be optimal for $j = 2 \bmod 4$.

Other useful facts for proving optimality are:

- (1) No sequence of $j > 4$ has $P = 0$ (Ref. 7).
- (2) No sequence of $j > 13$ and $j = 1 \bmod 4$ has $P = 1$.

$$(3) \quad \sum_{i=1}^j a_i = \left(\sum_{i=1}^j c_i \right)^2$$

This latter equation is easy to derive. Suppose c_i has h 1's and ℓ -1's where $h + \ell = j$. Then

$$\sum_{i=1}^j c_i = h - \ell$$

To find the autocorrelation sum, it is sufficient to realize that every element of the sequence will be multiplied by h 1's and ℓ -1's; thus,

$$\sum_{i=1}^j a_i = (h - \ell)^2$$

F. Examples of Optimality Proofs

To illustrate methods of optimality proofs, two examples are given

1. Proof for $j = 36$. The period $j = 0 \bmod 4$, so each a_i must be $0 \bmod 4$

$$\sum_{i=1}^{36} a_i = \left(\sum_{i=1}^{36} c_i \right)^2$$

The right hand side of the above equation is an even square (0, 4, 16, 36, 64, 100, ...). The left hand side is

$$36 + \sum_{i=1}^{35} a_i$$

Thus either

$$\left| \sum_{i=1}^{35} a_i \right| = 0$$

or

$$\left| \sum_{i=1}^{35} a_i \right| \geq 20$$

If

$$\left| \sum_{i=1}^{35} a_i \right| \geq 20$$

then $M \geq 80$. If

$$\left| \sum_{i=1}^{35} a_i \right| = 0$$

then either every $a_i = 0$, which is impossible, or some $|a_i| \geq 4$. In this case $a_{j-i} = a_i$ gives two equal deviations from zero (unless $j - i = i$) which can be negated only by two other elements of the autocorrelation. With 4 nonzero elements, $M \geq 64$. If $j - i = i$, then this deviation from zero must be negated by two or more elements of the autocorrelation; thus $M \geq 8^2 + 4^2 + 4^2 = 96$. So $M = 64$ is optimal.

2. **Proof for $j = 41$.** The period $j = 1 \pmod 4$, so each a_i must be $1 \pmod 4$.

$$\sum_{i=1}^{41} a_i = \left(\sum_{i=1}^{41} c_i \right)^2$$

The right hand side is (1, 9, 25, 49, 81, ...); therefore

$$41 + \sum_{i=1}^{40} a_i$$

equals the right hand side (RHS). If $\text{RHS} = 81$, $40 a_i = 1$ give $M = 40$ and $P = 1$, which is impossible. If $36 a_i = 1$, $2 a_i = -3$, and $2 a_i = 5$, then $M = 36 + 18 + 50 = 104$. If $\text{RHS} = 49$,

$32 a_i = 1$ and $8 a_i = -3$ give $M = 32 + 72 = 104$. If $\text{RHS} = 121$, $30 a_i = 1$ and $10 a_i = 5$ give $M = 30 + 250 = 280$. These are the minimum deviations from all $a_i = 1$ for the RHSs closest to 81. Thus $M = 104$ is optimal.

G. Further Evidence of Optimality

Evidence of optimality can also be obtained even when an exhaustive search has not been performed and a proof attempt indicates that lower values of M may be possible. When $j = 44$, for example, 2^{44} sequences are possible. However, as our discussion of equivalent sequences has shown, if one sequence has a given M value, so do a number of others:

44 rotations

$\times 2$ +/- interchanges

$\times 20$ modifications by taking every n th element where $(44, n) = 1$.

This appears to give 1760 sequences with the same autocorrelation properties. However, some of these sequences are identical, so the number of equivalent sequences is less than 1760. Nevertheless, for some values of M , more than 2^{10} equivalent sequences with a given M value must exist if any do. So roughly every 10 billionth sequence would have that value. If anywhere near that many sequences were examined iteratively without finding a given M value, it would strongly suggest that either no sequence giving that value existed or that such a sequence could not easily be derived by iterative one-bit modifications of better and better sequences.

When the M -value is underlined in Table 2, the authors feel that no better value exists. When no proof of optimality exists, the best evidence for this is the accumulation of a large number of sequences, many of which are equivalent, of that value. When one hundred sequences of $M = 144$ for $j = 44$ are found but none of $M < 144$ are discovered, there is considerable circumstantial evidence that $M = 144$ is optimal for $j = 44$. On the other hand, $M = 112$ for $j = 60$ may seem very surprising. Prior to the discovery of such a sequence, one might be excused for believing that no such sequence will be found. Yet, when one or two sequences with $M = 112$ are discovered, the evidence against a sequence with $j = 60$ and $M = 80$ or 96 is not overwhelming. Thus, the authors have decided not to underline a value for M in the table unless at least 30 sequences with that M have been discovered independently.

Of course, this criterion is no guarantee of optimality. For example, for $j = 43$, thirty sequences with $M = 138$ were discovered prior to the appearance of a lower value ($M = 42$). The generation of $P = 1$ sequences for $j = 43$ and $j = 47$ by the iterative method in less than 30 minutes of Cyber processing

time also indicates that the M -values listed for $j < 47$ are likely to be optimal.

H. Ratio of Ones to Minus Ones

As can be seen from the preceding section, the best sequences have

$$\left(\sum_{i=1}^j c_i \right)^2 \approx j$$

(the DC Fourier element)

$$\Rightarrow \sum_{i=1}^j c_i \approx \sqrt{j}$$

for best results. Thus one expects that the best codes of $j \approx 64$ will have

$$\sum_{i=1}^j c_i = \pm 8$$

For a period of 64, this gives 36 1's and 28 -1's (or vice versa).

Robert Keston has pointed out that if one splits such a $j = 0 \pmod 4$ sequence, c , into two sequences, f and g , where

$$\begin{aligned} f_i &= c_{2i-1} \\ g_i &= c_{2i} \end{aligned}$$

then either f or g should have an equal number of 1's and -1's (R. Keston, personal communication, May 1984).

This can provide assistance in selecting initial sequences or discarding unwanted sequences.

If one is looking for a *particular* value of M for a given j , it may be helpful to look at the properties of the a_i and c_i that must be satisfied. For example, at one time the best known M for $j = 48$ was 112. It was hoped that an M of 96 could be obtained. To accomplish this, one must have:

$$\begin{aligned} \sum_{i=1}^{48} a_i &= \left(\sum_{i=1}^{48} c_i \right)^2 \\ &= 16, 36, 64, 100, \dots \\ &= \text{RHS} \end{aligned}$$

But $|\text{RHS} - 48| > 24$ would mean $M > 96$ so only $\text{RHS} = 36$ or 64 are possible. For RHS to equal 36, there must be an odd number of fours in the autocorrelation; this can not give $M = 96$. Also, $\text{RHS} = 64$ cannot work with 6 fours, since 5 of them, including the middle element would be positive and one (the middle element again, which is impossible) would be negative. So the only sequence which works must have an autocorrelation with +8 in the middle and a +4 on either side so that $48 + 8 + 4 + 4 = 64$. The sequence itself must have either 20 or 28 1's. If one takes every other element of the sequence, one will get 8, 12, or 16 1's. These restrictions could make it easier to hunt for such a sequence. Luckily in this case, even without using them, there was ample time to find a sequence with $M = 96$.

I. Sequences With Good Matched but Poor Mismatched Autocorrelation Properties

Sequences with excellent mismatched autocorrelation properties generally have very good matched autocorrelation properties. The converse is not true and is most typically false for quadratic residue, double-prime or shift register sequences. The reason is that for such sequences,

$$\left| \sum_{i=1}^j c_i \right| = 1$$

So the first element of the Fourier transform,

$$\lambda_1 = \frac{1}{\sqrt{j}}$$

Suppose $j = 63$. Then

$$\lambda_1 = \frac{1}{\sqrt{63}}$$

$$\frac{1}{\lambda_1^2} = 63$$

Using Parseval's theorem:

$$\sum_{i=1}^{63} \lambda_i^2 = 63$$

Thus

$$\sum_{i=2}^{63} \lambda_i^2 = 63 - \frac{1}{63}$$

At best, all the λ_i for $2 \leq i \leq 63$ are equal. Then each of these

$$\lambda_i^2 = \frac{1}{62} \left(63 - \frac{1}{63} \right) = \frac{64}{63}$$

So

$$\sum_{i=1}^{63} \frac{1}{|\lambda_i|^2} = 63 + \frac{62 \times 63}{64} = \frac{63^2}{32}$$

Therefore the L value is, at best, $63/32 = 1.96875$, which is very poor. The L value of the shift register sequence with $j = 63$ and $M = 62$ in Table 2 actually is $63/32$; the remaining Fourier components are equal.

By the above argument the best possible L value for a shift register sequence of $j = 2^n - 1$ is:

$$L = \frac{2^n - 1}{2^{n-1}}$$

Unless n is small, $L \approx 2$, or about 3 dB.

The DC component of the Fourier transform, greatly elevated due to the near equality of 1's and -1's in the sequence, always produces an L value which represents roughly a 3 dB loss in signal; this compares very unfavorably with the 0.15 dB to 0.25 dB losses corresponding to some of the sequences with better ratios of 1's to -1's. Sequences of period $4j$ formed by multiplying an $L = 1$ sequence of period 4 by a shift register sequence are no better, as the L -value of the sequence with period $4j$ equals that of the sequence with period j .

This provides another incentive for not investigating shift register sequences exclusively. Not only is it a nuisance to analyze such sequences; in addition their autocorrelation properties are, in some respects, not very good.

John Bailey has offered a solution to this problem; eliminate the DC component. One method would be to have +1 and -1 be out of phase by other than 180° (Ref. 8).

For a two phase sequence the modification is:

Element of Shift-Register Sequence	Element of Modified Sequence
+1	+1
-1	- exp $i\beta$

where

$$\beta = \tan^{-1} \left(\frac{2\sqrt{j}}{j-1} \right)$$

For example, if the unmodified sequence is

$$+++--+-$$

then the modified sequence is

$$c = (1, 1, 1, -\exp i\beta, -\exp i\beta, 1, -\exp i\beta)$$

For complex elements, the autocorrelation is:

$$a_\ell = \sum_{k=1}^j c_k^* c_{k+\ell}$$

In our example, $a_7 = 7$

$$\begin{aligned} a_\ell &= 3 - 2(\exp i\beta) - 2(\exp -i\beta) \\ &= 3 - 4 \cos \beta \end{aligned}$$

for $1 \leq \ell \leq 6$ where

$$\beta = \tan^{-1} \frac{\sqrt{7}}{3}$$

$$\tan^2 \beta + 1 = \frac{1}{\cos^2 \beta} \Rightarrow \cos^2 \beta = \frac{9}{16}$$

Putting β in the first quadrant, $\cos \beta = 3/4$, so $a_\ell = 0$ for $1 \leq \ell \leq 6$.

In general, when $j = 2^N - 1$

$$a_\ell = \frac{j-1}{2} - \frac{j+1}{2} \cos \beta$$

for $1 \leq \ell < j$. Thus

$$\tan^2 \beta = \frac{4j}{(j-1)^2}$$

and

$$\cos^2 \beta = \frac{(j-1)^2}{(j+1)^2}$$

which, with β in the first quadrant, gives

$$a_q = 0 \quad (q \neq 0 \text{ mod } j)$$

One could also derive a 3-phase sequence as a product of two 2-phase sequences:

Elements of Unmodified Sequences		Element of Modified Product
1	1	1
-1	-1	1
1	-1	$-\exp(-i\beta)$
-1	1	$-\exp i\beta$

Once again,

$$\beta = \tan^{-1} \left(\frac{2\sqrt{j}}{j-1} \right)$$

If one wishes to zero all sidelobes without a decrease in SNR, one can also let $j = 2^{2N}$ and create a sequence with \sqrt{j} phases; a complete discussion of this would be too far afield of the topic of binary sequences.

V. Results

Table 2 shows the best sequences for periods 28 to 64 for both matched and mismatched cases. Table 1, showing the results for periods 3 to 27 (Ref. 9) is included for completeness.

In Table 2, the heading j gives the period (length) of the sequence; P gives the lowest value of the peak sidelobe; M gives the lowest sum of the squares of the sidelobes discovered for any sequence of period j . When the sequence with the optimal peak sidelobe has a higher M , both values are given. When two references are given on the same line, the first one refers to the matched sequence and the second one to the mismatched sequence. A reference of "X" refers to this article.

When the value for P , M or L is in parentheses, the authors feel that a better, but as yet undiscovered, sequence may exist. When the value is underlined, it is unlikely that a better value exists. In all other cases, the value can be proved to be optimal. All values for P are optimal unless two values are given for a specific j , in which case the lower one is optimal.

The sequences are written in hex notation. The first bit is always a plus sign. For example, the sequence for 29 is given in hex as 14A7C111. In binary this would be 0001 0100 1010 0111 1100 0001 0001 0001. By replacing 0 with a minus sign, and 1 with a plus sign, and removing the leading zeros, we get the sequence:

+ - + - - + - + - - + + + + - - - - + - - - + - - - +

References

1. Boehmer, A., Binary Pulse Compression Codes, *IEEE Transactions on Information Theory*, Volume 13, No. 2, p. 156, April 1967.
2. Lindner, J., Binary Sequences up to Length 40 with Best Possible Autocorrelation Functions, *Electronic Letters*, Vol. II, No. 21, p. 507, 16 October 1975.
3. MacMullen, A., *Radar Antennas, Transmitters, and Receivers*, pp. PC-1-56, Technology Service Corporation, April 1977.
4. Hoffman de Visme, G., *Binary Sequences*, English Universities Press, 1971.
5. Golomb, S., Baumert, L., Easterling, M., Stiffler, J., and Viterbi, A., *Digital Communications With Space Applications*, Prentice-Hall, 1964.
6. Brennan, L. E., and Reed, I. S., Theory of Adaptive Radars, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-9, March 1973.
7. Turyn, R., *Optimum Codes Study*, Sylvania Electronic Systems Final Report AF19 (604)-5473, 29 January 1960.
8. Bailey, J., *Modified PN Codes With Optimum Autocorrelations*, Technology Service Corporation, August 1978.
9. Tyler, S., and Keston, R., Optimal Periodic Binary Codes of Lengths 28 to 64, *TDA Progress Report 42-57*, March and April 1980, Jet Propulsion Laboratory, Pasadena, CA.
10. Bailey, J., and Tyler, S., *Periodic Binary Waveforms with Optimum Autocorrelation Functions*, Technology Service Corporation, September 1978.
11. Tausworthe, R., *Correlation Properties of Cyclic Sequences*, JPL Technical Report No. 32-388, 1 July 1963, Jet Propulsion Laboratory, Pasadena, CA.
12. Watkins, J., Lofjsson, J., and Tyler, S., A Binary Sequence of Period 60 with Better Autocorrelation Properties than the Barker Sequence of Periodic 13. *TDA Progress Report 42-82*, April-June 1985, Jet Propulsion Laboratory, Pasadena, CA.
13. Barlog, M., Horn, T., Ulrich, D., and Variot, M., *Binary String Manipulation*, CSUN Report CS480, May 1985.

Table 1. Best matched and mismatched sequences for periods 3 to 27

| Matched Sequence | | | | Mismatched Sequence | |
|------------------|----------|----------|---------------|---------------------|---------------|
| <i>j</i> | <i>P</i> | <i>M</i> | Sequence, hex | <i>L</i> | Sequence, hex |
| 3 | 1 | 2 | 4 | 1.5000 | 4 |
| 4 | 0 | 0 | E | 1.0000 | E |
| 5 | 1 | 4 | 1E | 1.1111 | 1E |
| 6 | 2 | 20 | 25 | 1.3125 | 28 |
| 7 | 1 | 6 | 4B | 1.5400 | 40 |
| 8 | 4 | 16 | CB | 1.3333 | E5 |
| 9 | 3 | 24 | 1F4 | 1.6650 | 104 |
| 10 | 2 | 36 | 350 | 1.6761 | 25D |
| 11 | 1 | 10 | 716 | 1.2909 | 67A |
| 12 | 4 | 16 | 941 | 1.1250 | 941 |
| 13 | 1 | 12 | 1F35 | 1.0400 | 1E6B |
| 14 | 2 | 52 | 36A3 | 1.2153 | 27F5 |
| 15 | 1 | 14 | 647A | 1.1520 | 698F |
| 16 | 4 | 48 | FAC4 | 1.2589 | EED8 |
| 17 | 3 | 64 | 19A3D | 1.2165 | 1128E |
| 18 | 2 | 68 | 31EDD | 1.2843 | 21419 |
| 19 | 1 | 18 | 7A86C | 1.1119 | 465D0 |
| 20 | 4 | 64 | F6E8E | 1.1111 | C5640 |
| 21 | 3 | 52 | 117BCE | 1.1097 | 170848 |
| 22 | 2 | 84 | 3D1231 | 1.2178 | 28312B |
| 23 | 1 | 22 | 6650FA | 1.1114 | 7CEA2D |
| 24 | 4 | 32 | DC20D4 | 1.0607 | C3DEA6 |
| 25 | 3 | 72 | 18B082E | 1.1195 | 128C0BC |
| 26 | 2 | 100 | 2C1AEB1 | 1.1240 | 34AFBC9 |
| 27 | 3 | 74 | 5A3C444 | 1.0965 | 7D3472B |

Table 2. Best matched and mismatched sequences for periods 28 to 64

| Matched Sequence | | | | Mismatched Sequence | | |
|------------------|----------|------------|------------------|---------------------|------------------|------------------------|
| <i>j</i> | <i>P</i> | <i>M</i> | Sequence, hex | <i>L</i> | Sequence, hex | Reference ^a |
| 28 | 4 | <u>80</u> | B30FDD4 | <u>1.1305</u> | F3DDD21 | 9, 10 |
| 29 | 3 | <u>92</u> | 14A7C111 | <u>1.1384</u> | 14A7C111 | 9 |
| 30 | 2 | 116 | 3FAD938A | <u>1.1260</u> | 3FAD938A | X |
| 31 | 1 | 30 | 4B3E3750 | <u>1.0898</u> | 45C2D660 | 11, 10 |
| 32 | 4 | <u>80</u> | 89445BC1 | <u>1.0950</u> | 89445BC1 | 10 |
| 33 | 3 | 64 | 18A5C240D | <u>1.0656</u> | 18A5C240D | 9 |
| 34 | 2 | 132 | 29D3BB82D | <u>1.1337</u> | 29D3BB82D | 10 |
| 35 | 1 | 34 | 71F721592 | (1.0732) | 722B92F3E | 11, X |
| 36 | 4 | 64 | F397A6517 | <u>1.0455</u> | F397A6517 | 9 |
| 37 | 3 | 84 | 1BD623E3B6 | <u>1.0771</u> | 1BD623E3B6 | 9 |
| 38 | 2 | 148 | 302162B8B6 | <u>1.1085</u> | 302162B8B6 | 10 |
| 39 | 3 | <u>86</u> | 60CD4F47BE | <u>1.0528</u> | 60CD4F47BE | X |
| 40 | 4 | <u>80</u> | DB9EAE05BC | <u>1.0575</u> | DB9EAE05BC | 9 |
| 41 | 3 | 104 | 1079731045A | <u>1.0723</u> | 1079731045A | 10 |
| 42 | 2 | 164 | 2CF51397B7C | (1.0523) | 2CF51397B7C | X |
| 43 | 1 | 42 | 653BE2E08D6 | (1.0786) | 5189822FC34 | 11, X |
| 44 | 4 | <u>144</u> | AFDE8AF8665 | <u>1.1022</u> | AFDE8AF8665 | 9 |
| 45 | 3 | <u>124</u> | 17473C9BFAD0 | <u>1.0667</u> | 17473C9BFAD0 | 9 |
| 46 | 2 | 180 | 2A2818CDBC16 | (1.0842) | 2A2818CDBC16 | X |
| 47 | 1 | 46 | 421A8D93A9EF | (1.0727) | 795220A780EC | 11, Y |
| 48 | (8) | <u>96</u> | 99803C312AB6 | <u>1.0375</u> | 99803C312AB6 | X |
| 48 | 4 | (112) | CBF089223A51 | | | 9 |
| 49 | 3 | (144) | 1C0B504676CB0 | (1.0799) | 1C0B504676CB0 | X |
| 50 | 2 | 196 | 236D4FF70651E | (1.0984) | 236D4FF70651E | X |
| 51 | 3 | (146) | 6D5DECF8433E8 | (1.0704) | 6D5DECF8433E8 | X |
| 52 | 4 | (128) | FDEE871D85B44 | <u>1.0400</u> | E8BEDC6F7AE24 | X |
| 53 | 3 | (164) | 11CAA3E46F7B65 | (1.0706) | 11CAA3E46F7B65 | X |
| 54 | 2 | 212 | 3917B588A2C302 | <u>1.0826</u> | 3917B588A2C302 | X |
| 55 | 3 | (214) | 7BCFB32717D0A5 | (1.0837) | 7BCFB32717D0A5 | X |
| 55 | (5) | (182) | 7F0AA13316DC34 | | | X |
| 56 | 4 | (208) | 852659EBA181B8 | (1.0993) | 852659EBA181B8 | X |
| 57 | 3 | (184) | 16A38C8BC7FD1AD | (1.0637) | 16A38C8BC7FD1AD | X |
| 58 | 2 | 228 | 3B64AAF8FDCE520 | (1.0896) | 3B64AAF8FDCE520 | X |
| 59 | 1 | 58 | 5D49DE7C1846D44 | (1.1003) | 6CF43BE8A12CF9D | 11, Z |
| 60 | 4 | (112) | EC757781362D6F9 | (1.0352) | EC757781362D6F9 | 12 |
| 61 | 3 | (204) | 1481F734DC7EEA74 | (1.0624) | 1481F734DC7EEA74 | X |
| 62 | 2 | 244 | 225746DC62583D20 | <u>1.0638</u> | 225746DC62583D20 | 9 |
| 63 | 1 | 62 | 4314F4725BB357E0 | (1.0830) | 408AB703D6597390 | 11, 13 |
| 64 | 4 | (240) | B24FEAE7E4529CF0 | (1.0538) | CD9BFF0E16D2AB98 | ZZ, X |

^a X Refers to this article.

Y T. Safer, personal communication, April 1985.

Z N. Lee, personal communication, October 1984.

ZZ J. Watkins, personal communication, June 1985.

A Binary Sequence of Period 60 With Better Autocorrelation Properties Than the Barker Sequence of Period 13

J. Watkins
 TRW, Redondo Beach, California

J. Loftsson and S. Tyler
 Reliability Engineering Section

A binary sequence of period 60 has been discovered which in some respects has better autocorrelation properties than the Barker sequence of period 13. When both sequences are processed using appropriate sidelobe-eliminating mismatched filters, the Barker sequence's main lobe is reduced by a factor of 1.040 or 0.17 dB, while the new sequence's main lobe is reduced by a factor of only 1.035 or 0.15 dB. This sequence is the first counterexample known to the authors of the hypothesis that the autocorrelation properties of all sequences of periods greater than 13 are inferior to those of the Barker period-13 sequence. Sequences of this type are very useful in radar and deep space communications, especially in situations where there is an adverse signal to noise ratio.

I. Introduction

A binary sequence is a string of bits. It can be thought of as a vector C where each C_i is a plus one or a minus one. A periodic sequence is one which is continuously repeated; for a binary sequence of period j :

$$C_{i+j} = C_i$$

for all i . The autocorrelation, a , of a sequence, C , of period j is:

$$a_i = \sum_{k=1}^j C_k C_{k+i}$$

The element a_j is called the main lobe; the remaining a_i are sidelobes. For example, if $C = (1, 1, 1, -1)$, $a = (0, 0, 0, 4)$. In this example, the sidelobes are zero. For a sequence to have good autocorrelation properties the sidelobes should be small.

The example used is perfect, but for sequence periods other than 4, the autocorrelation sidelobes are never all zeros.

The cross-correlation, X , of two sequences, B and C , each of period j is

$$X_i = \sum_{K=1}^j B_K C_{K+i}$$

Let C be a binary sequence which is to be cross-correlated with B , a sequence composed of real numbers. The sequences C and B are related by the weighting function, T .

$$B_i = T_i C_i$$

In this case, B can be considered a "mismatched filter" to C .

For a mismatched filter B to be normalized,

$$\sum_{i=1}^j B_i^2 = j$$

Starting with a binary sequence, for example;

$$C_1 = (1, -1, 1, -1, -1, -1, -1, -1, 1, 1, 1, -1, -1, 1, -1, -1)$$

and the (normalized) weighting function,

$$T_i = \sqrt{2/11} (1, 4, 1, 2, 1, 4, 1, 2, 1, 4, 1, 2, 1, 4, 1, 2)$$

then the normalized mismatched filter

$$B_i = \sqrt{2/11} \times (1, -4, 1, -2, -1, -4, -1, -2, 1, 4, 1, -2, -1, 4, -1, -2)$$

and the cross-correlation

$$X_i = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \sqrt{2048/11})$$

Examples of periodic binary sequences with good autocorrelation properties are available from a number of sources as shown in the companion paper to this article (Ref. 1). Such sequences are utilized extensively in satellite and planetary probe ranging and communications (Ref. 2). Mismatched weighting functions can often suppress the cross-correlation sidelobes to arbitrarily low levels, but only at the expense of a loss in signal to noise ratio (SNR) (Ref. 3).

This loss equals the square of the sequence period divided by the square of the main lobe. In the above example of a sequence of period 16 analyzed with a normalized mismatched filter to give no sidelobes, the loss is:

$$\text{loss} = 256 \times \frac{11}{2048} = \frac{11}{8} = 1.375$$

The loss in dB is $10 \log_{10} (1.375) = 1.38$ dB. A method for calculating sidelobe-eliminating weighting functions is given in Ref. 1.

A periodic sequence is a Barker sequence if its sidelobes are all +1. A Barker sequence can have two periods, 5:

++++-

and 13:

++++- - +- +- +

The Barker sequence of period 13 has a reduction in main lobe of a factor of 1.040 when processed by a mismatched filter which completely suppresses all cross-correlation sidelobes, corresponding to a loss in signal to noise ratio of 0.17 dB (Ref. 4). Until now, the authors had not been able to discover a counterexample to the hypothesis that no sequence of period greater than 13 will have better autocorrelation properties than those of this Barker sequence. For example, the best codes listed for periods 36, 39, and 64 all have somewhat greater SNR losses (0.19, 0.22, and 0.23 dB, respectively) than does the Barker code of period 13.

II. A New Sequence of Period 60

A search of binary sequences of period 60 was performed on a Cyber 750 computer. The Cyber was chosen due to its availability, speed, and the inclusion in its assembly language of an instruction which counts the number of ones in a word. This latter property greatly facilitates a search for sequences which have good matched-filter autocorrelation properties (Ref. 5). The period was chosen to correspond to the Cyber's word length of 60. Good sequences were found iteratively by looking at all one-bit modifications of the previous best sequence.

The sequence evaluator, written in COMPASS, was decisive in permitting an adequate number of sequences to be evaluated in a reasonable amount of time. The program segment in Fig. 1 was used for this task.

The Cyber processed approximately 10^8 sequences per hour. Within 50 computer-hours, two sequences were discovered each having an excellent matched-filter autocorrelation. One of them is:

```

+ + - + - + - - + - + - - - + + - + - + - - - - - + -
- - - + - + - + - - + + + - - - + - + + - - - + - - - +

```

Represented in hexadecimal notation, the sequence is:

D651B4021678B91

The other sequence, in hex notation, is:

EC757781362D6F9

By starting with bit 52 of the second sequence and taking every 53rd bit modulo 60, one obtains the ones complement of the first sequence. Both sequences have SNR losses of only 0.15 dB when analyzed by the appropriate mismatched filter.

III. Implications for Aperiodic Sequences

An aperiodic sequence does not repeat. For it, the autocorrelation, a , of a sequence, C , of length ℓ is

$$a_i = \sum_{K=1}^{\ell-i} C_K C_{K+i}$$

The main lobe is a_0 ; the remaining a_i are sidelobes.

An aperiodic sequence is a Barker sequence if the absolute values of each of its sidelobes never exceeds 1. Aperiodic Barker sequences exist for all sequence lengths less than 6 as well as for lengths 7, 11, and 13.

The question of whether the aperiodic Barker sequence of length 13 is superior to aperiodic sequences of all other lengths has not yet been completely resolved. M. Golay has defined the "merit factor" for aperiodic sequences to be the square of the sequence length divided by twice the sum of the squares of the sidelobes (Ref. 6). His original upper bound for the merit factor was $2e^2 = 14.778$, which is approached only by the Barker-13 sequence's merit factor of 14.08. Golay states that 12.325 is a reasonable upper bound on the merit factor of all other aperiodic sequences. He has performed an exhaustive search of skew symmetric sequences of lengths up to 59 to support this. Whether or not the discovery of the present periodic sequence of length 60 stimulates further tests of Golay's bound, it should help to dispel the notion that the autocorrelation properties of the best periodic as well as aperiodic sequences degrade with increasing length.

References

1. Tyler, S., and Loftsson, J., Periodic binary sequences with very good autocorrelation properties, *TDA Progress Report 42-82*, Jet Propulsion Laboratory, Pasadena, CA, August 1985.
2. Golomb, S., Baumert, L., Easterling, M., Stiffler, J., and Viterbi, A., *Digital Communications With Space Applications*. New York: Prentice-Hall, 1964.
3. Ackroyd, M., and Ghani, F., Optimum mismatched filters for sidelobe suppression, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-9, pp. 214-217, March 1973.
4. Rihaczek, A. W., and Golden, R. M., Range sidelobe suppression for Barker codes, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-7, pp. 1087-1092, Nov. 1971.
5. Corning, B., Heraper, D., Lee, R., Rothman, M., and Watkins, J., Binary sequences of periods 60, 64, and 72, CSUN Report CS380, California State University, Northridge, CA, May 1984.
6. Golay, M., The merit factor of long low autocorrelation binary sequences, *IEEE Transactions on Information Theory*, Vol. IT-28, pp. 543-549, May 1982.

| | | | |
|------|-------|---------|---------------------------------------|
| | IDENT | MVAL | Call assembler with ID |
| | ENTRY | MVAL | Define entry point |
| MVAL | EQ | \$+1S17 | |
| | SB2 | 30 | Put number of shifts into B2 |
| | SB1 | 1 | Used for decrementing |
| | SX5 | 30 | For (number of 1's - 30)*2 |
| | SX6 | B0 | Sum of squares of sidelobe |
| | SA2 | X1 | Put parameter value in X2 |
| | BX3 | X2 | Put sequence to be shifted in X3 |
| LOOP | LX3 | 1B | Left rotate the sequence by one bit |
| | BX4 | X3-X2 | XOR the shifted and original sequence |
| | CX4 | X4 | Count the ones in the result |
| | IX4 | X4-X5 | Subtract 30 from the number of 1's |
| | IX4 | X4+X4 | Double the difference |
| | IX4 | X4*X4 | Square the difference |
| | IX6 | X6+X4 | Add this to the total |
| | SB2 | B2-B1 | Decrement loop counter |
| | NE | B2,LOOP | Loop if not equal to 0 |
| | IX6 | X6+X6 | Double the sum |
| | IX6 | X6-X4 | Avoid doubling middle term |
| | EQ | MVAL | This returns |
| | END | | |

Fig. 1. Sequence evaluator

The GCF Mark IV Implementation and Beyond

R. A. Crowe
TDA Engineering Section

This article presents a brief overview of the major subsystems that comprise the Mark IVA GCF along with key information on basic functionality that may be inadequately or imperfectly understood. Concluding paragraphs describe the evolving GCF System as it is currently being designed.

I. Introduction

The Ground Communication Facility (GCF) was modified in order to meet the requirements of the DSN Mark IVA implementation now nearing completion. The key characteristics of the GCF did not change markedly during this time frame (see Figs. 1-3). The underlying architecture of the GCF was left intact while modifications were initiated in order to expand the capacities of its constituent subsystems. Communications support at Goldstone, Spain and Australia was unified and consolidated under the aegis of the incipient Deep Space Communications Complex (DSCC). Economics of shared functionality have evolved whereby the collective communications capacity at each DSCC is shared by the competing demands of its DSS components. The DSN was expanded in this time frame to begin assuming responsibility for the scheduling and operation of the 26-meter network. One 26-meter antenna is located at each of the three DSCC sites. Voice, data and teletype circuits required for Spain and Australia continue to be ordered from the NASA Communications (NASCOM) network. NASCOM Engineering has delegated to the GCF the overall responsibility for communications support between Goldstone and JPL. This has resulted in expansion of the intersite Goldstone microwave facilities that link DSCC 10 with the Communications Center known as GCF 10, and the

installation of a fiber-optic communication path between the DSCC and the outlying 26-meter station.

Current plans call for continued GCF expansion in the area of capacity and augur significant changes in the area of key characteristics. High-rate missions and Very Long Baseline Interferometry (VLBI) tracks late this decade and into the 1990s provide the principal motivations toward added capability. Efficiency, generality, automation and simplification militate toward a restructuring of the communication architecture to embody layering techniques, variable size data units, and interfaces to external networks.

II. Digital Communications Subsystem

The Digital Communications Subsystem performs the exchanges of digital data blocks between the DSCC's and the Central Communication Terminal (CCT) at JPL (Fig. 4). Block multiplexing on the available communication circuits permits shared use of the lines among projects. Communication capability at each DSCC has been implemented in a modular fashion in order to save on circuit lease costs. Four configurations are possible at each DSCC:

- (1) One 56 kilobits/sec (kb/s) duplex circuit

- (2) One 56 kb/s duplex circuit and one 56 kb/s simplex circuit
- (3) One 56 kb/s duplex circuit and two 56 kb/s simplex circuits
- (4) One 56 kb/s duplex circuit and one 224 kb/s simplex circuit.

All simplex circuits carry data in the direction of the DSCC at JPL only. Configuration changes are possible by scheduling/descheduling of circuits. No software changes are required.

The Digital Communications Subsystem is composed of the following principal assemblies:

A. Station Digital Communication (SDC) and Central Digital Communication (CDC) Subsystems

The SDC and CDC Subsystems consist of all data transmission equipment including line interfaces, data sets, Network Encoders Decoders (NED), Communication Buffers (CB), digital and analog test equipment, and patch facilities for trouble isolation as well as a front-end line interface between the actual data sets and NED appearances (see Figs. 5 and 6).

The data transmission equipment is provided by NASCOM including the line interfaces/converters, data sets and common carrier interfaces associated with the overseas DSCCs.

The CDC implementation includes a 40-by-40 digital line switch that provides the connection from the data set to a particular NED/CB combination that plugs in to a specific port on one of the five Error Correction and Switching (ECS) front-end computers at the CCT. This switch expedites normal activation of data circuits and simplifies the switchover to the backup ECS in the event of hardware failure.

B. Area Routing Assembly (ARA)

The ARA is the DSN computer component that performs the GCF functions at the DSCC. Two identical Mod Comp II/25 computers are so configured: one for prime and one for backup. The ARA receives data blocks from all DSCC computers and multiplexes them to JPL on the available circuitry. In the reverse direction, data received from JPL is distributed to DSCC computers according to destination code and data type.

The ARA in cooperation with the ECS will perform error correction of data blocks by retransmission of data transferred on the 56-kb/s duplex circuit. The ARA will attempt retransmission of an unacknowledged block one time only.

The ARA accepts and transmits either 1200-bit or 4800-bit data blocks on the available circuitry. The circuits are utilized as follows:

- (1) All 1200-bit data blocks will be sent on the 56-kb/s duplex line. They cannot be sent on the simplex line(s).
- (2) All 4800-bit data blocks are normally sent on the simplex circuits. When necessary and if loading permits, all 4800-bit data traffic from a specified link (a link is usually equivalent to a DSS) may be moved from the simplex line to the duplex line.

Based on these facts and on the configuration possibilities discussed earlier, the following conclusions are possible:

- (1) While the maximum GCF DSCC to JPL rate is 280 kb/s (56 + 224), no single link-originated 4800-bit stream may exceed 224 kb/s, and no single 1200-bit stream may exceed 56 kb/s.
- (2) Error correction is normally applied to the low rate 1200-bit traffic from the DSCC.

In addition to these functions, the ARA writes Original Data Record (ODR) tape files of data transmitted. The maximum recording rate is restricted to the maximum transmission rate: 280 kb/s. Data from each DSS is normally recorded on its own tape drive although data from multiple DSSs can be combined on a single tape if desired.

C. Error Correction and Switching (ECS) Computer

The ECS is the DSN computer component resident at the JPL CCT that interfaces the communications circuits from the DSCCs. There are five ECS computers configured identically for support. Four serve a prime data handling function and one is designated as a backup. Responsibilities include servicing serial interfaces to Remote Mission Operation Centers (RMOC), Remote Information Centers (RIC), the Mission Control and Computing Center (MCCC) and the Network Operations Control Center (NOCC).

Capabilities of the ECS include the following:

- (1) Error Correction of data blocks in cooperation with ARA at the DCCCs
- (2) Multiple routing of data blocks to the NOCC, GCF Data Records and to any two other processing centers
- (3) Delivery of data blocks predicated on the source originating the data
- (4) A multilevel real-time dump facility enabling full traceability of data blocks through the program.

III. Analog Intersite Communications

The Mark IVA implementation at Goldstone left the DSS 12 antenna and front-end equipment in place approximately 13 miles from the Signal Processing Center (SPC) at the DSCC. In addition, the NASCOM STDN 26-meter antenna (DSS 16) at the Goldstone site was itself about 7-8 miles from the SPC of the DSCC. In order to transfer baseband signals to/from these two antenna sites, it was necessary to expand and enhance the existing Intersite Communications Subsystem. Microwave links were used to connect DSS 12 to SPC-10, and a fiber-optics link was installed to provide the baseband conduit between DSS 16 and SPC-10.

The existing microwave channels between DSS 12 and SPC-10 were expanded from 8 to 16 duplex channels. A channel is here defined as a communications link in one direction. The following are provided:

- (1) Two receiver baseband channels
- (2) Two command modulation channels and one confirmation channel
- (3) One full duplex frequency and timing channel
- (4) One frequency STD channel
- (5) One FEA-12 surveillance TV channel
- (6) Two test modulation (SIM) channels
- (7) Two full duplex diversity multiplex channels that support voice, 56-kb/s wideband circuits, LAN, gateway circuits, 230.4 kb/s simplex circuit and low speed and control circuits.

At DSS 16 it was necessary to connect the front-end equipment at the receiver with the telemetry and command equipment at SPC-10 in order to provide DSN/NASCOM cross support capability for projects beginning with AMPTE. This capability was provided with the completion in July 1984 of a six-fiber underground installation of a fiber optics system with optical transmitters and receivers located both at DSS 16 and at SPC-10. At the present time, up to 3 baseband signals may be transmitted in each direction by multiplexing FM modulated carriers over the fiber optic link. Currently, there are two fibers in use for telemetry and command baseband signal transfer, with additional usage planned for voice and ultimately digital transfers when the 26-meter network is fully integrated into the DSN.

IV. Data Records Subsystem

The Data Records Subsystem is comprised chiefly of the three prime Data Records Generator (DRG) computers and

one backup computer resident at the CCT. Functional components of this subsystem reside also at the DSCC in order to provide gap "fill" information for data not received in real-time. The principal function of the DRGs is to create tape Intermediate Data Record (IDR) files for project delivery.

The DRG software checks selected data streams for correct source code, spacecraft I.D., User Data Type, time tag, block serial number and error status code. The DRG detects gaps in the data and outputs real-time statistics providing the percentage of good data received. At the end of a pass, a recall operation may be initiated requesting data from the DSCC in order to improve the percent of good data on the IDR.

V. Voice Subsystem

The Mark IVA voice assemblies at the DSCCs were configured from the tactical intercom assemblies (TICS), Communications junction modules, and station switch assemblies that were previously in use. There were no changes made to the equipment at the JPL CCT end (Fig. 8). At the Signal Processing Center (SPC) of the DSCC, new communication panels were developed for use at the front console. New circuit boards were developed to bring the obsolescent TIC panels more in line with current technology.

VI. Implementation Activities

A DSN system level approach is being taken to redesign the GCF to meet the communication necessities of the 1990s. Both Local Area Networks and wide area networks will form part of this system. A software layering architecture shall be employed in these networks in order to simplify user interfaces to the system and allow modification to lower level structures without affecting the upper echelon layers. Key changes during this implementation include megabit/second communication rates from the DSCCs, automatic recall of data *before* the tracking pass is complete, processing and handling of variable size data blocks from the user and Gateway interfaces to external networks located both at the DSCC and the CCT. Most of these upgrades will be operational by 1989.

A. High-Rate DSCC Communication

The current maximum data rate from each DSCC will be increased from the present 280 kb/s to 1,544 kb/s. The design calls for this capability to be modular so that the Communications capacity at any DSCC may be incrementally adjusted to fit the combined requirements of all active projects in any given time frame. Error correction of data blocks will be extended to encompass the entire bandwidth of data transmitted from the DSCC.

B. Near Real-Time Recall

The spare bandwidth available on the DSCC communication circuits shall be utilized in order to allow recall of data blocks before the tracking pass is complete. By this method, the data that could not be successfully transmitted in real-time because of line outages or transient overloads will be shunted to a temporary data file. This data will subsequently be retransmitted when loading and line conditions allow on the excess available bandwidth from the DSCC to JPL. Independent of any operator intervention, the CCT data records function will organize the playback data into separate files and merge the blocks with the original real-time stream in order to provide the most complete data file.

C. Variable Size Data Blocks

Perhaps the most revolutionary innovation to the GCF System will be the capability of accepting variable size data units from all originating computers. Within specified mini-

mum and maximum length constraints, the data originator shall be free to deliver to the receiver data blocks that are sized to match perfectly the application being supported. This will contribute to overall network efficiency and minimize the overhead labor required to prepare data for transmission. This is also consistent with packetization.

D. Gateway Interfaces

Gateway interfaces will be developed at the DSCC and at the CCT in order to permit data communication with networks external to the DSN. The architecture for gateway interfacing will be installed when the GCF system is upgraded. Actual gateway software will be developed upon identification of the requirement to exchange data with a particular network. The location of gateway interfaces at the overseas sites will permit circuit economies ensuing from the proximity of the external network to the DSN access point at the DSCC.

References

1. Evans, R. H., "DSN Ground Communications Facility," *TDA Progress Report 42-65*, pp. 9-18, July and August 1981, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1981.
2. Mortensen, L. O., "GCF Mark IV Development," *TDA Progress Report 42-70*, pp. 161-172, May and June 1982, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1982.
3. Crowe, R. A., "Design Issues in the GCF Mark IV Development," *TDA Progress Report 42-75*, pp. 132-139, July and September 1983, Jet Propulsion Laboratory, Pasadena, Calif., November 15, 1983.

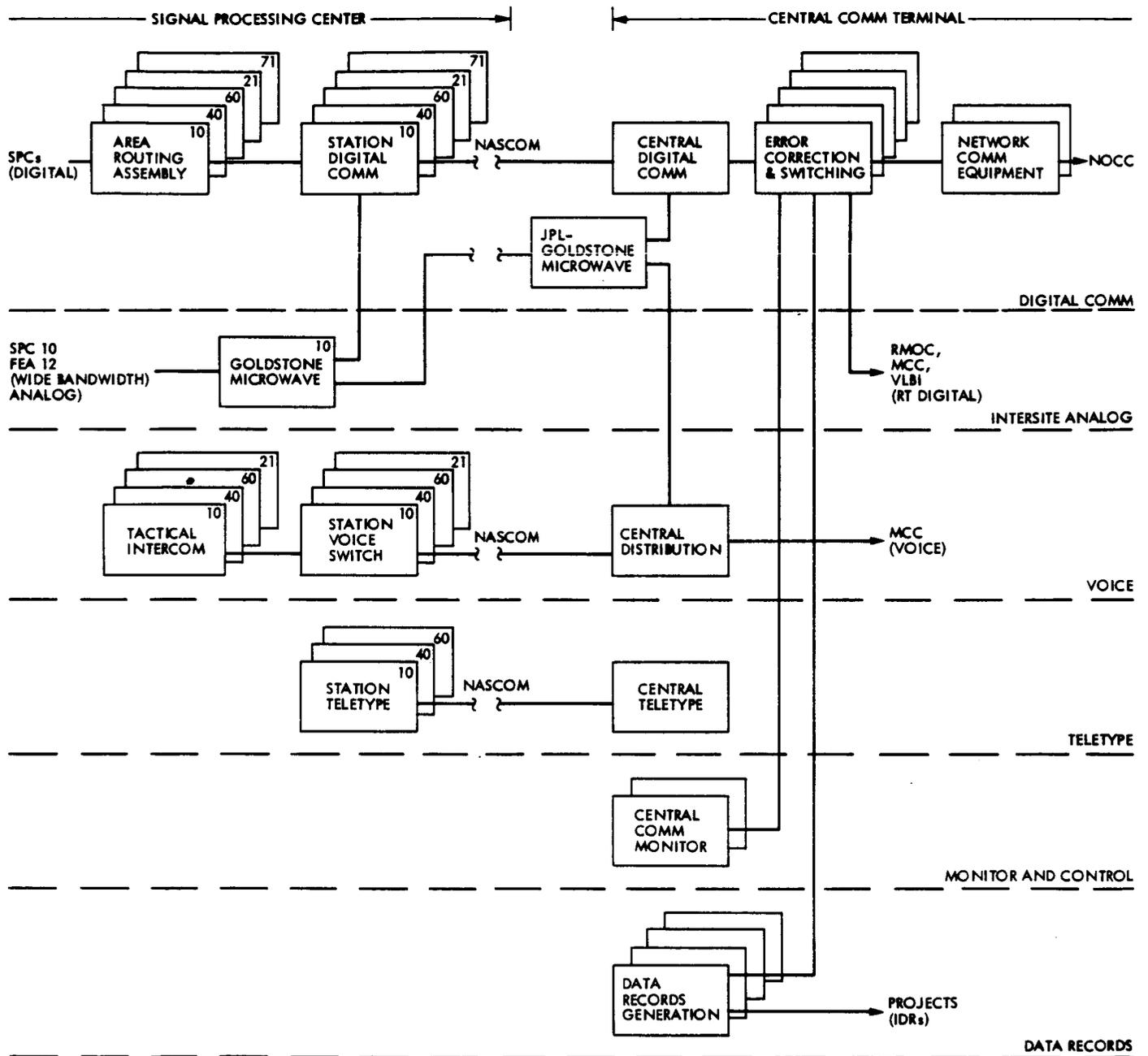


Fig. 1. Mark IV-A era GCF subsystems

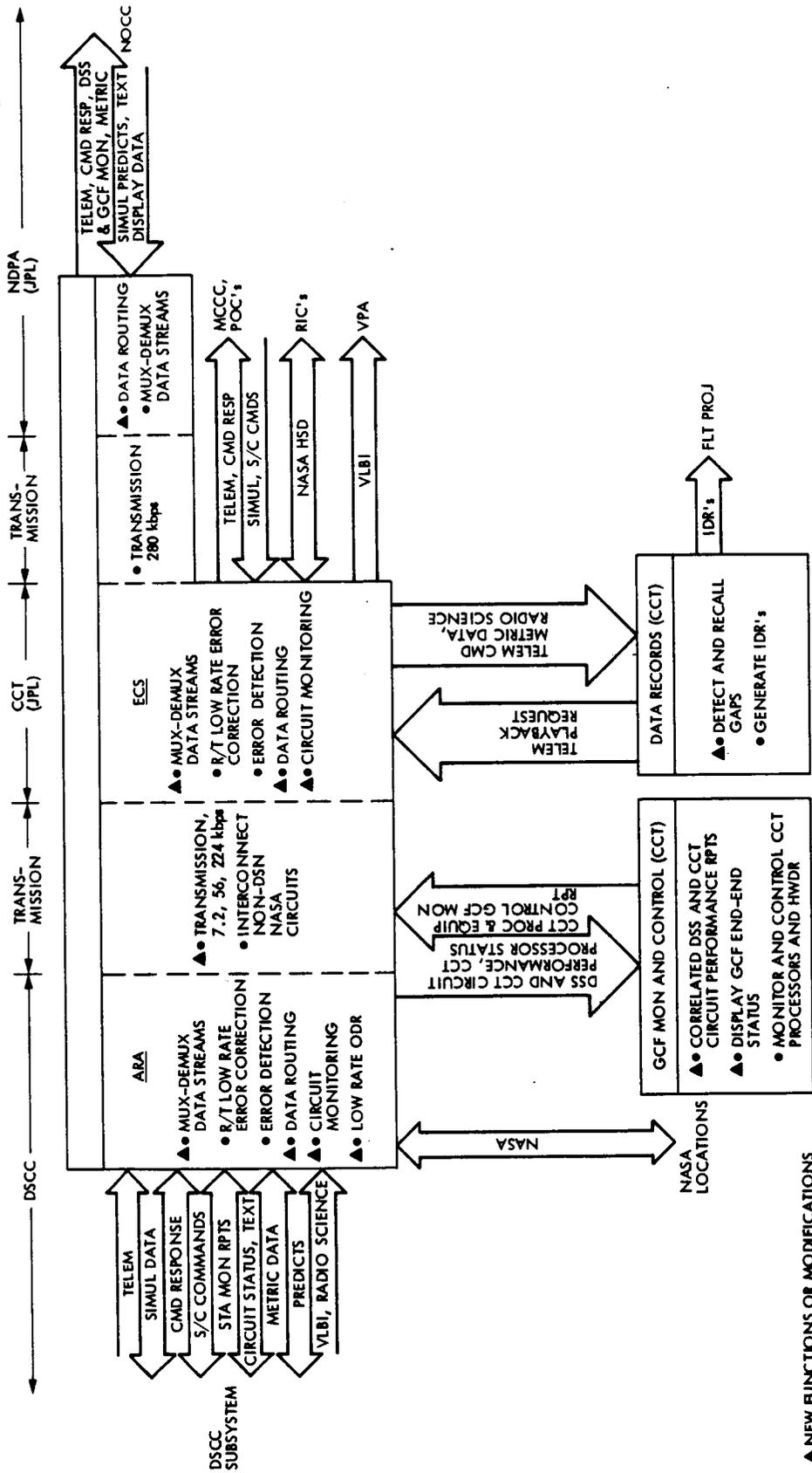


Fig. 2. Mark IV-A end-to-end digital communications, data record, monitoring, and network communications equipment functions and interfaces

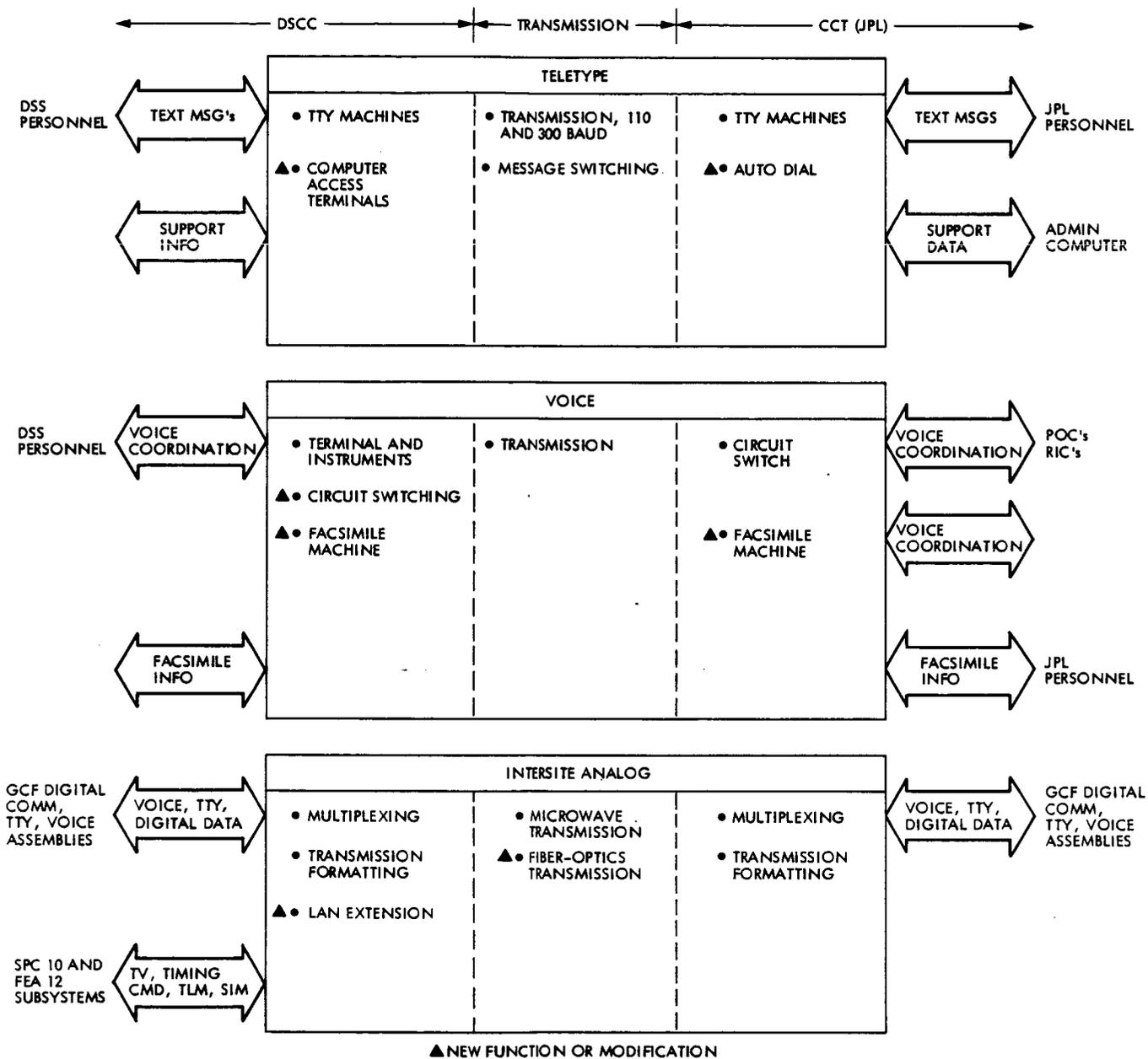


Fig. 3. Mark IV-A end-to-end teletype, voice, and intersite analog functions and interfaces

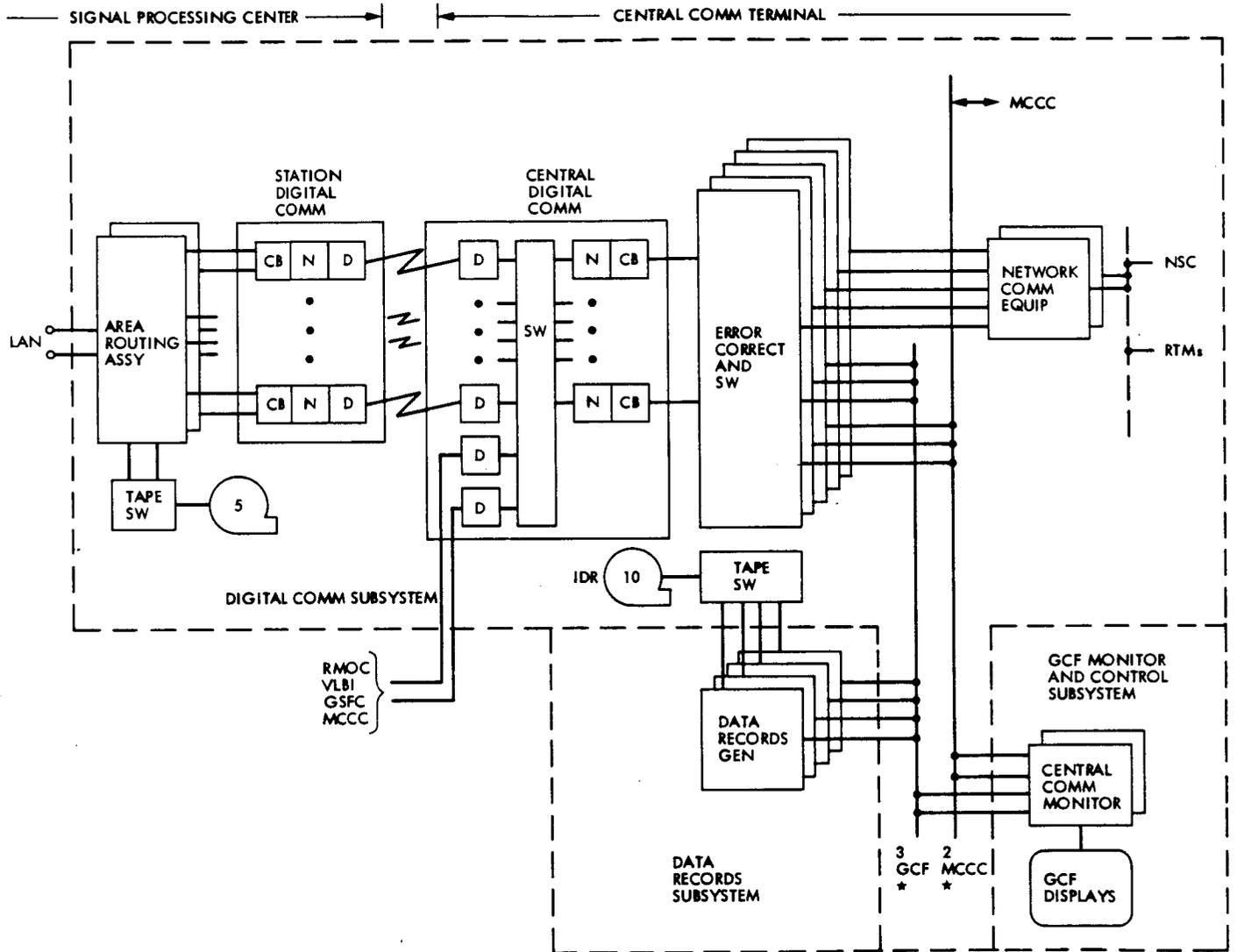


Fig. 4. Digital communications subsystem overview of GCF digital configuration

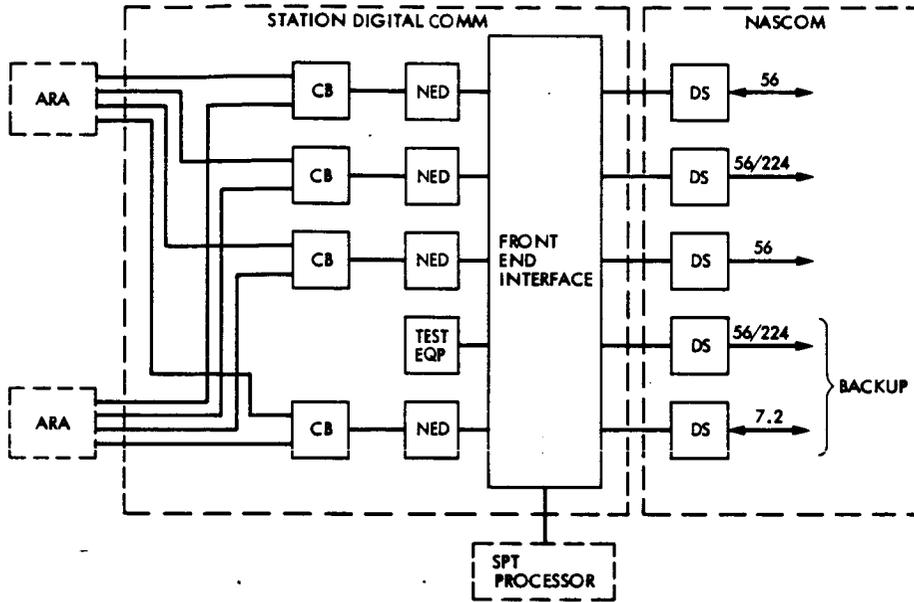


Fig. 5. Data transmission, station digital communications

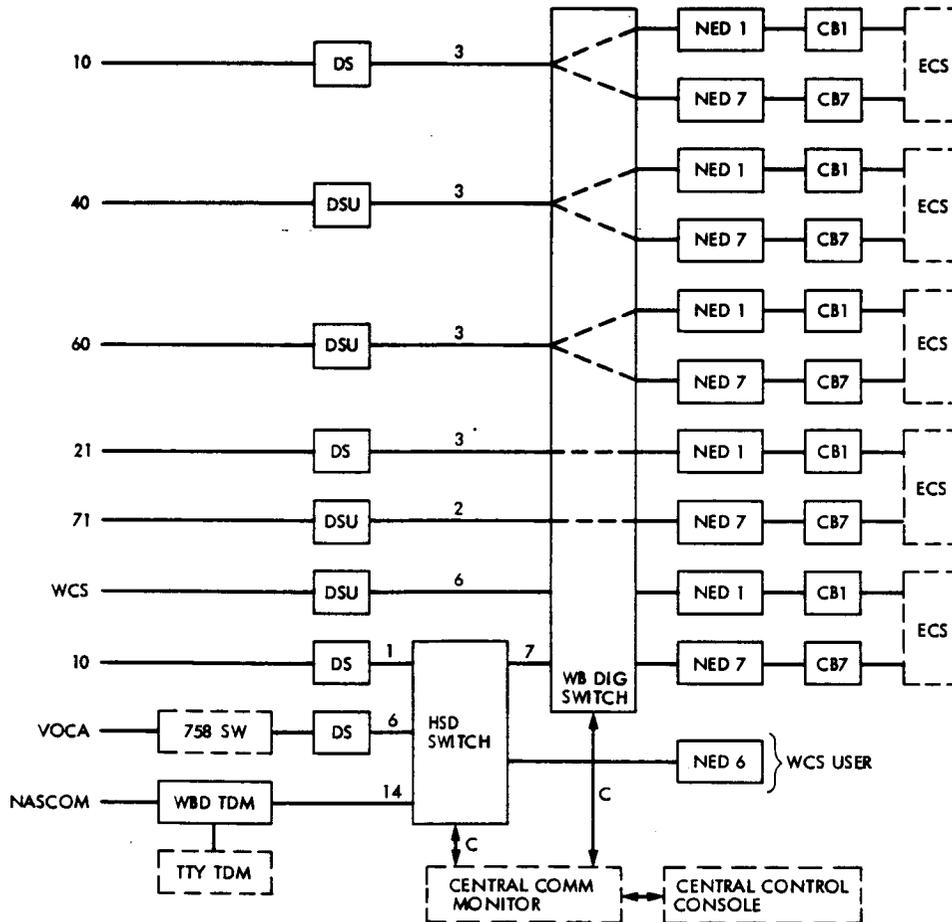


Fig. 6. Data transmission, central digital communications

C-3

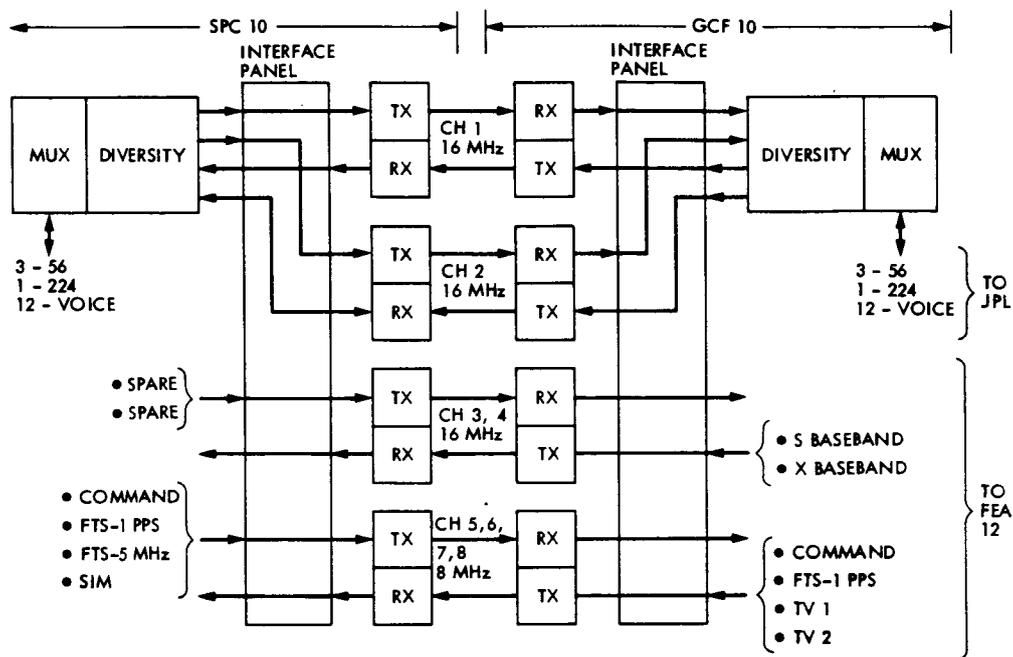
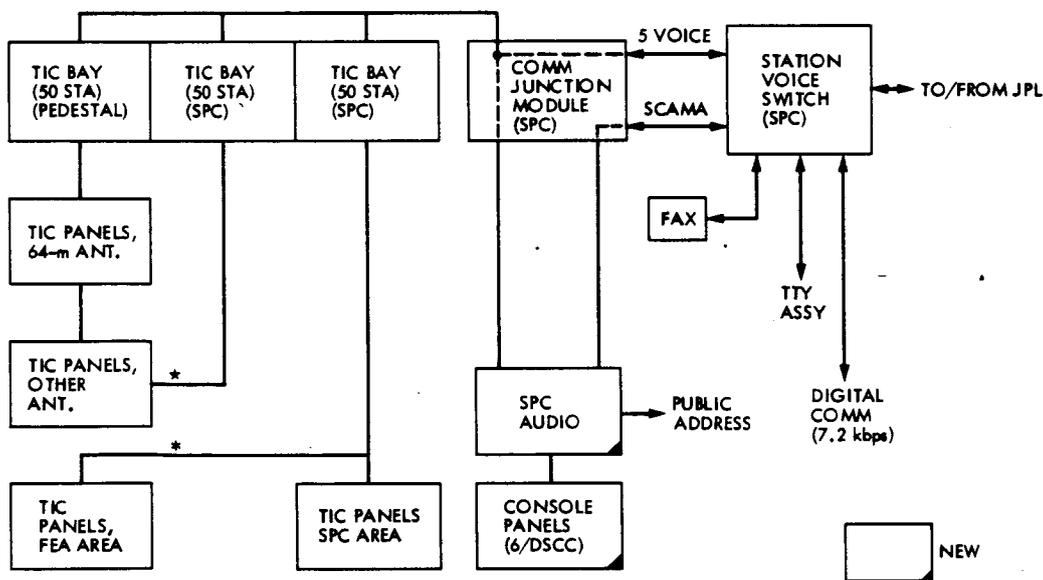


Fig. 7. Intersite microwave configuration, general design



- * WILL BE MICROWAVED IN THE CASE OF FEA-12
- + SCAMA SERVICE AVAILABLE AT OVERSEAS CONSOLES ONLY
- * MAY REQUIRE AMPLIFICATION/POWER FOR PANELS IN DISTANT ANTENNAS

Fig. 8. DSCC voice, general design

SETI Radio Spectrum Surveillance System

B. Crow

Communications Systems Research Section

A. Lokshin

Space Physics and Astrophysics Section

M. Marina and L. Ching

Radio Frequency and Microwave Subsystems Section

The SETI Radio Spectrum Surveillance System (SRSSS) will provide a data base for assessing the RFI environment for SETI and minimizing RFI disruptions during the search. This article describes the system's hardware and software and discusses the sensitivity of the system.

I. Introduction

The Search for Extraterrestrial Intelligence (SETI) Radio Spectrum Surveillance System (SRSSS) was built to provide a data base to assess the RFI environment for SETI. Initially the system will be used to study the RFI environment at the Goldstone Complex. A survey will be conducted over the next year to cover the frequency range from 1 to 10 GHz. The SRSSS is sufficiently portable so that it can be moved easily to other radio astronomical sites (see Figs. 1 and 2). Later, the SRSSS will be used as a coincidence detector with the SETI search instrument.

The SRSSS was designed as a stand alone system capable of detecting and recording RFI signals in an unattended mode. Periodic system calibration is achieved using a noise diode to provide a known input power. Detected signals are stored on floppy diskettes and will be processed with the SETI/Radio Astronomy VAX 11/750 at the Jet Propulsion Laboratory (JPL).

The SRSSS data base will provide guidance to the SETI project on how to search the frequency, time and space dimensions. The objective is to minimize the probability of RFI disrupting the search.

This article gives an overall description of the system, including hardware and software, and a discussion of the system sensitivity.

II. Hardware

The SRSSS occupies one standard rack of equipment and has a one-meter antenna, antenna rotator and associated RF-module assembly mounted externally. The rack contains a Tektronix 494p Spectrum Analyzer under the control of a Tektronix 4052A controller. The controller also controls the antenna position, amplifier selection and data storage on a dual-floppy disk. A hard copy unit is available to record "interesting" spectra. The data from the floppy disk will be

processed in a VAX computer at JPL. An overall block diagram is presented in Fig. 3.

A. RF Module

The signal from the antenna is amplified in one of seven GaAsFET amplifiers. The set of seven GaAsFET amplifiers is required to cover the spectrum from 1 to 10 GHz. Specifications of the GaAsFET amplifiers are given in Table 1.

Each GaAsFET amplifier is followed by a transistor amplifier with a minimum gain of 20 dB to eliminate subsequent cable loss from contributing to the system noise temperature. The measured gain and noise temperature for the RF assembly are shown in Figs. 4 and 5.

The selection of different sets of amplifiers for the specific frequency ranges is controlled by the 4052A Tektronix computer through HP actuators. A directional coupler is used at the input of the coaxial switch to inject noise into the system for the purpose of system calibration.

Calibration of the system is performed by turning the HP346B noise source on and off and calculating the system temperature from the y -factor measurement in the 4052 controller. This noise source has a 15.2 dB \pm 0.2 dB excess noise ratio (ENR) over a range of 10 MHz to 18 GHz. The low SWR of the noise source in both the on and off condition reduces a major source of measurement uncertainty. The ENR is related to the effective noise temperature of the source (T_{ne}):

$$ENR = 10 \log \frac{(T_{ne}-290)}{290}$$

B. Relay Actuators

The HP relay actuators provide the control of the microwave coaxial switch and the HP 346B noise source via control lines to the RF unit located in the antenna rotator pedestal.

C. Antenna Control

The antenna elevation is manually set at a fixed elevation angle (adjustable from 0 to 30 degrees). The azimuth position and rate of rotation are controlled by the RPM antenna controller through the Tektronix 4052A computer via an IEEE-488 Bus.

The azimuth position of the antenna may be controlled in local mode with the front panel control. The rate of rotation is displayed in RPM on the RPM LED indicator. The rotator angular position is simultaneously displayed on the azimuth DEGREES LED indicator. In a remote controlled operation,

rate or position command signals from the IEEE-488 Bus may be used to control the rate of rotation of the rotator from -15.5 RPM to +15.5 RPM in 0.5 RPM increments, or to cause the rotator to point at any azimuth with a 0.1 degree resolution and 0.15 degree accuracy.

D. Spectrum Analyzer

The 494P Tektronix Programmable Spectrum Analyzer is controlled by the 4052A Tektronix Computer through the IEEE-488 Bus to select the desired frequency range, frequency span/division and resolution bandwidth.

This spectrum analyzer has a frequency range of 10 KHz to 21 GHz and a minimum resolution bandwidth of 30 Hz. The sensitivity is -121 dBm at 30 Hz resolution bandwidth.

E. Digital Clock

The HP 59309A digital clock supplies the system time. This clock can be set under local control or by remote commands received from the IEEE-488 Bus.

F. Dual Disk Drives

The double density dual floppy diskette storage facilities can hold up to two megabytes of information which translates to 50,000 messages (assuming 40 characters per message). These messages typically will contain the following data: time, frequency, azimuth, threshold, number of hits above threshold, operating mode, etc. At low threshold setting, which may result in 1,000 messages per hour, the SRSSS can operate unattended for a full weekend. At a more conservative threshold, say 60 messages per hour, the SRSSS can operate unattended for a full month.

G. Hard Copy Unit

The 4631 Tektronix Hard Copy Unit provides hard copies of any spectra displayed on the 4052A CRT. A sample display is shown in Fig. 6.

III. Software

The software that runs on the SRSSS is written in Tektronix BASIC. The program allows a user to build and save a customized observation program on the cartridge tape. This schedule will later direct the system when and where to point the antenna, how to set the threshold, frequency range and resolution bandwidth by issuing detailed hardware commands.

There are five main modules in the SRSSS software. The EDITOR allows the operator to build and edit observation schedules. The RUN module runs observation according to

the prerecorded schedule. The MAINTENANCE and UTILITIES modules aid in software development and hardware diagnostics. Figure 7 is the top level flow diagram of the SRSSS. The RUN module is further expanded in Figs. 8, 9, and 10. The EDIT module flow diagram is shown in Fig. 11.

An observation schedule is a set of tasks (up to 20). The tasks of a schedule are separated into 10 Sequential Time Events (STE) tasks and 10 Absolute Time Event (ATE) tasks.

The STE tasks are executed by the SRSSS system one after another. An ATE task starts and stops according to its predetermined absolute time schedule, and it interrupts any currently running STE task. An interrupted STE task resumes execution as soon as the interrupting ATE task is completed.

Any task, whether an STE task or ATE task, falls into one of three categories:

1. Data Only
2. Calibration Only
3. Data and Calibration

In a Data Only task, the SRSSS determines and records hits above some predetermined threshold. In a Calibration Only task, system calibration operations are performed. In a Data and Calibration task, the SRSSS interleaves calibration operations with hits collected. Figure 12 shows schematically the

azimuth-frequency coverage for a DATA/CALIBRATION task.

The SRSSS was intended for unattended data collection, but its software provides some convenient functions for an observer to get information in real time. By request, the system can work in a stepped, frame-by-frame mode, reporting hits on the CRT, plotting current spectra on the screen and making hard copy.

The SRSSS software utilizes the fact that the spectrum analyzer contains an internal controller which allows some parallel processing. This allows the spectrum analyzer to collect the data, while the 4052 controller processes that data from the last spectrum.

IV. Future Plans

Currently, the JPL SETI team is working on building data base software on the VAX 11/750 for RFI signal storage and analysis. This data base will provide an estimate of how many RFI events may be expected for a given SETI run. Based on this information an optimal pattern of sky/frequency coverage in SETI SKY Survey will be developed. A year long survey is planned to assess the RFI environment at Goldstone. This RFI data base should be of general interest for all potential DSN users.

Table 1. Specifications of the GaAsFET amplifiers

| Amplifier | Frequency Range, GHz | Minimum Gain, dB | Noise Figure, K |
|-----------|----------------------|------------------|-----------------|
| AR1 | 1.1 - 1.7 | 44.5 | 100 |
| AR2 | 1.7 - 2.3 | 37 | 100 |
| AR3 | 2.3 - 3.1 | 38 | 120 |
| AR4 | 3.1 - 4.3 | 36 | 120 |
| AR5 | 4.3 - 5.8 | 40 | 160 |
| AR6 | 5.8 - 7.7 | 42 | 225 |
| AR7 | 7.7 - 10.0 | 37 | 390 |

ORIGINAL PAGE IS
OF POOR QUALITY

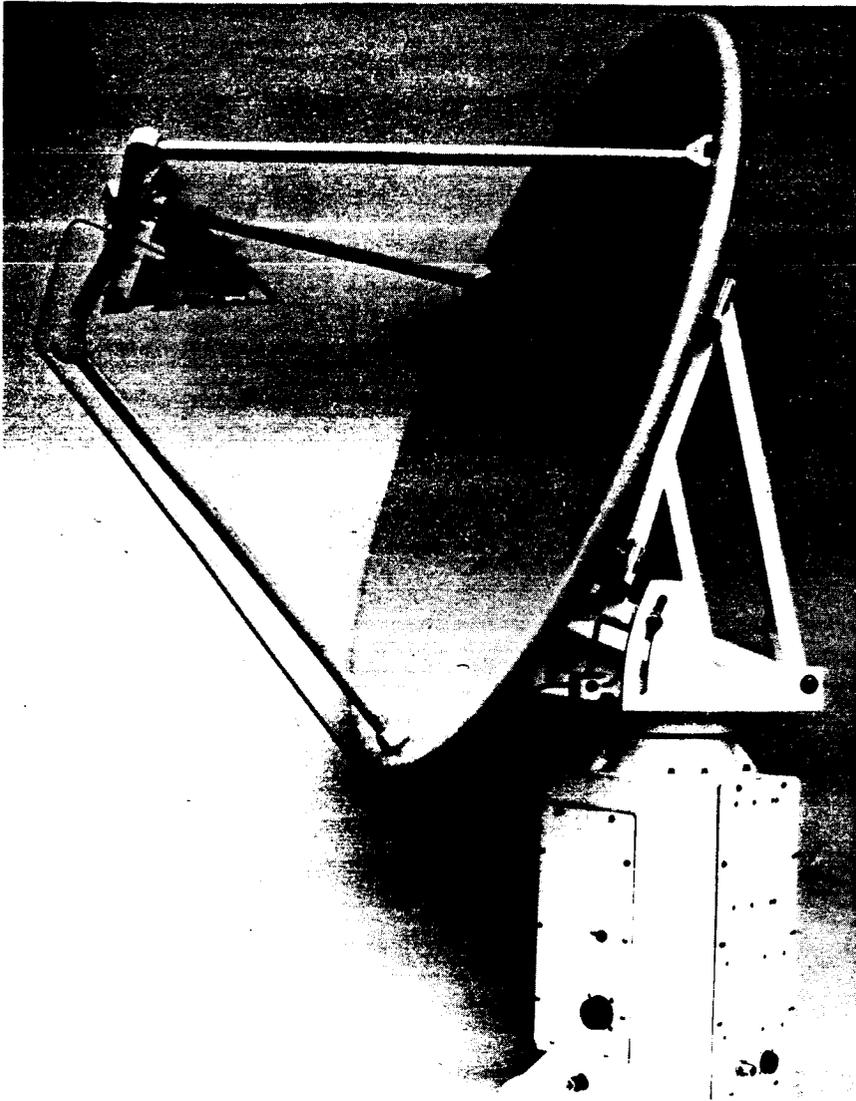


Fig. 1. SRSSS antenna assembly

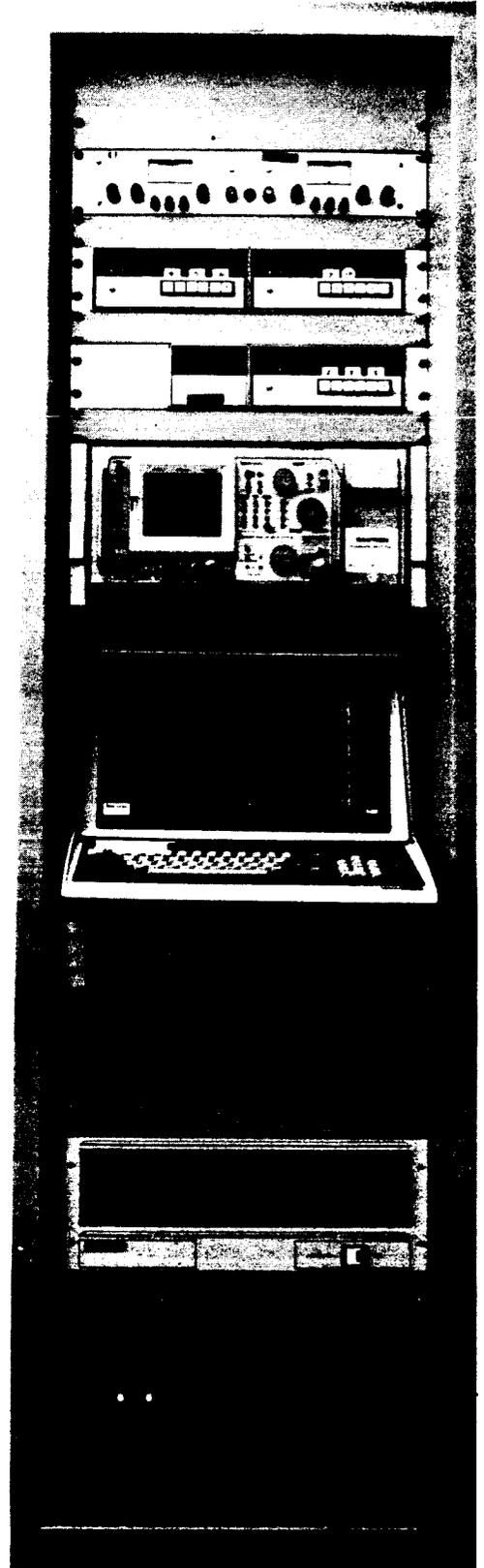


Fig. 2. SRSSS electronics assembly

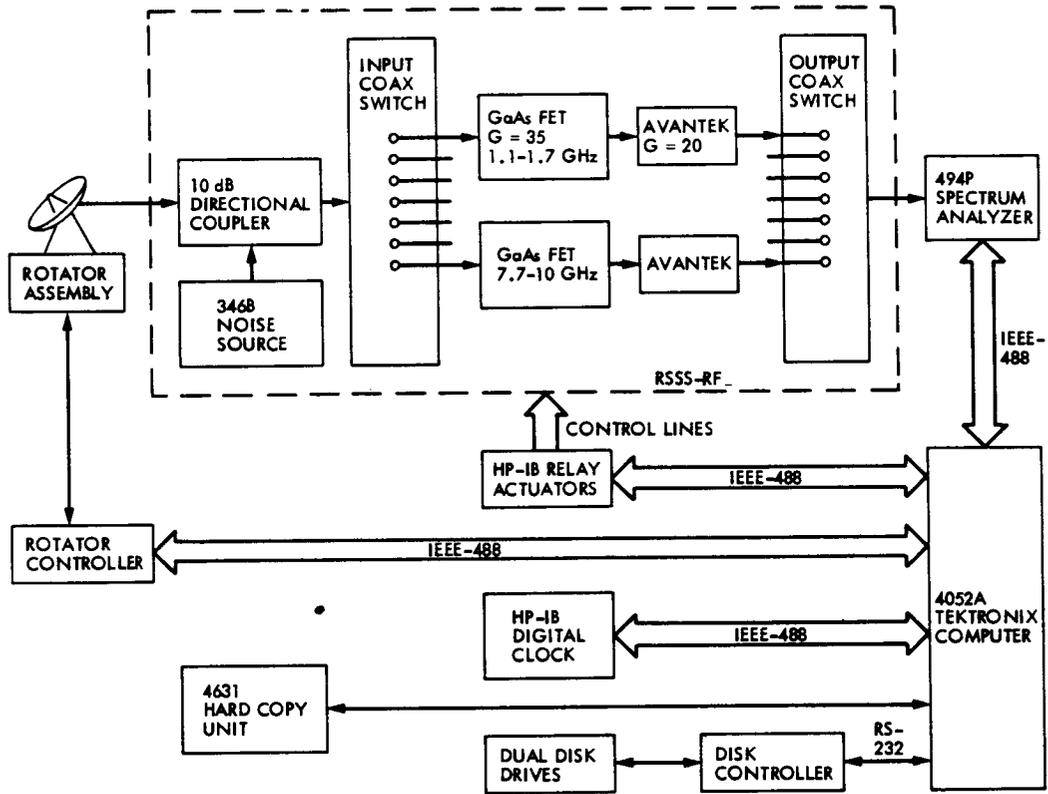


Fig. 3. Radio Spectrum Surveillance System (RSSS) block diagram

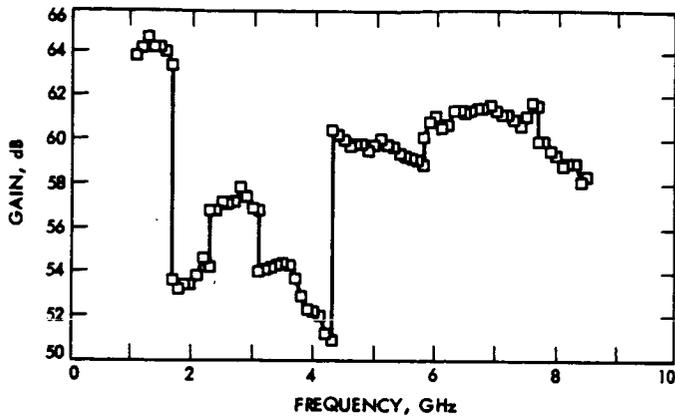


Fig. 4. Gain vs frequency for the SRS3 amplifier

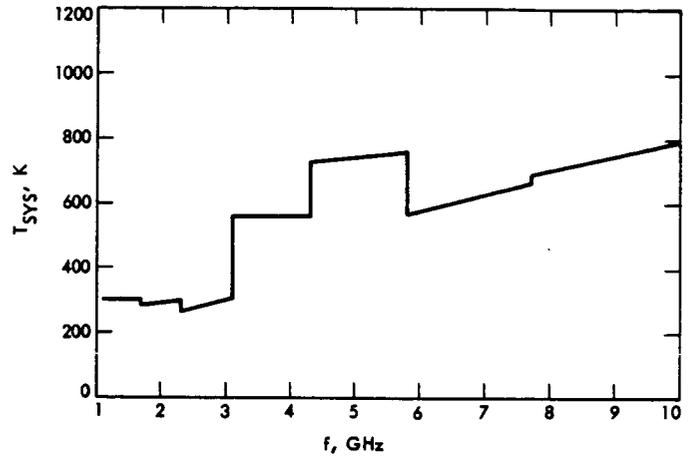


Fig. 5. The SRSSS system temperature vs frequency, April 1985

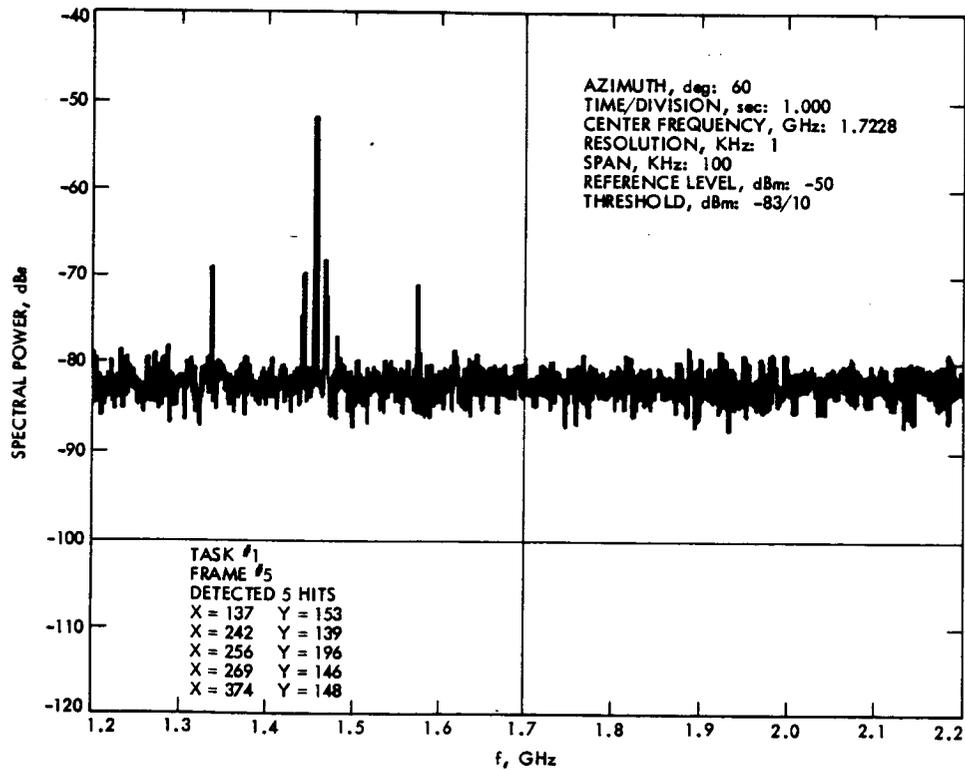


Fig. 6. The SETI Radio Spectrum Surveillance System

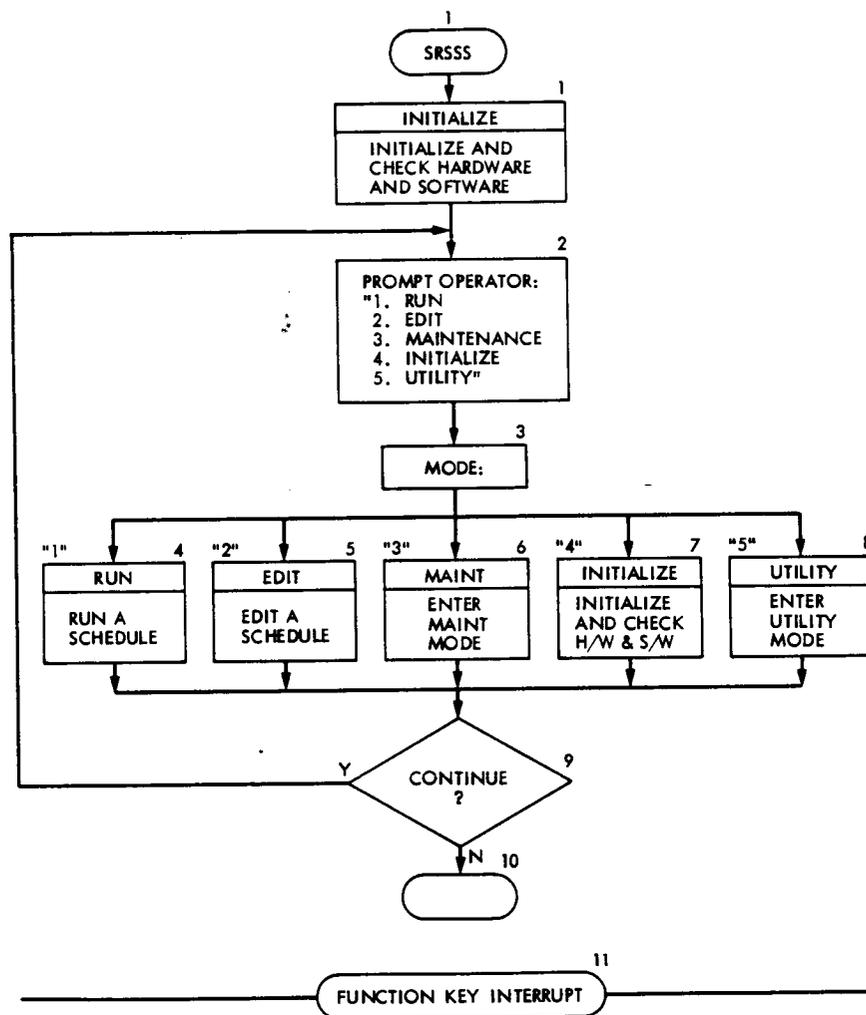


Fig. 7. The main program of SRSS

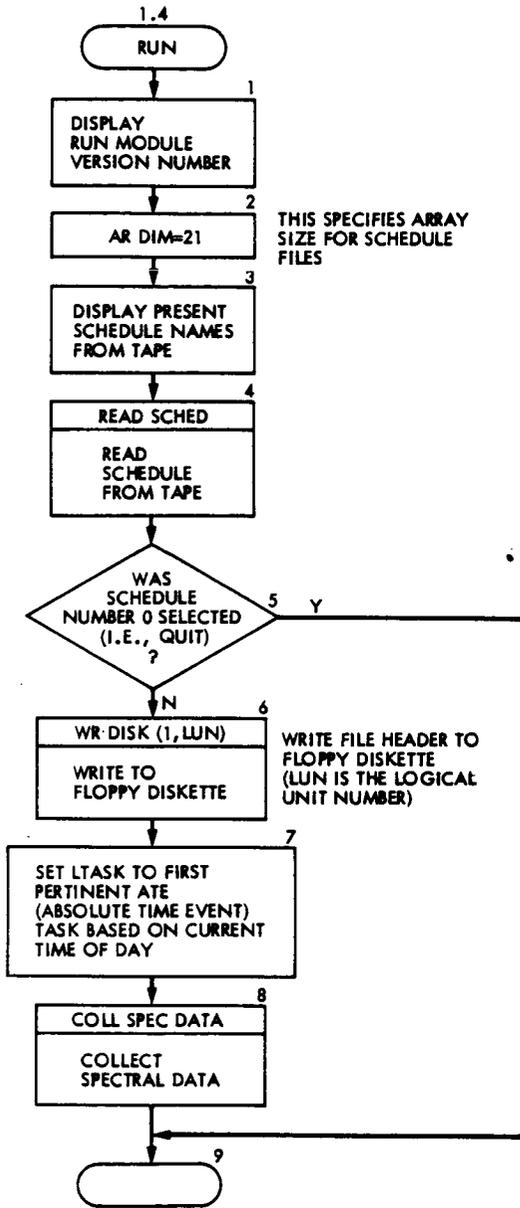


Fig. 8. RUN mode flowchart

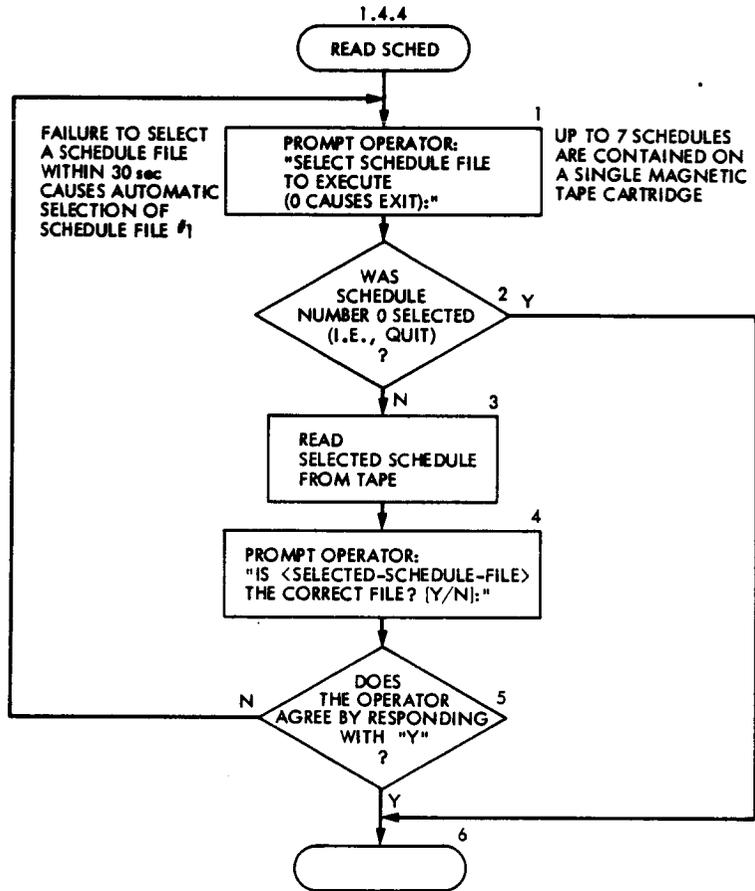


Fig. 9. READ schedule flowchart

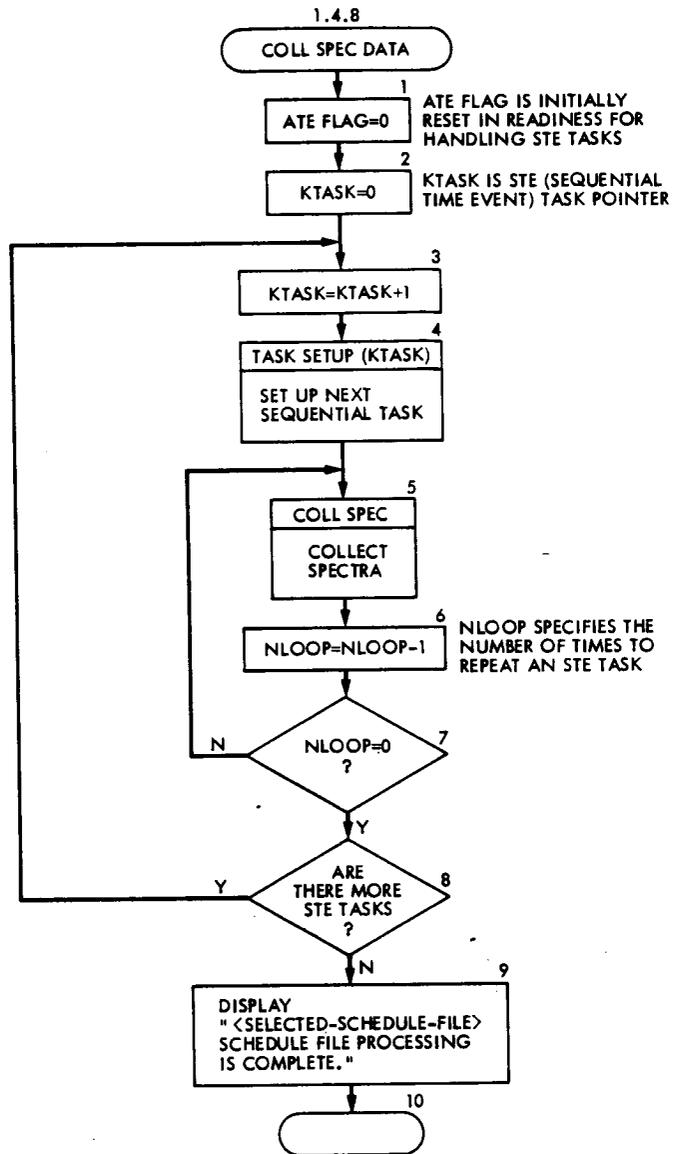


Fig. 10. Collect spectral data flowchart

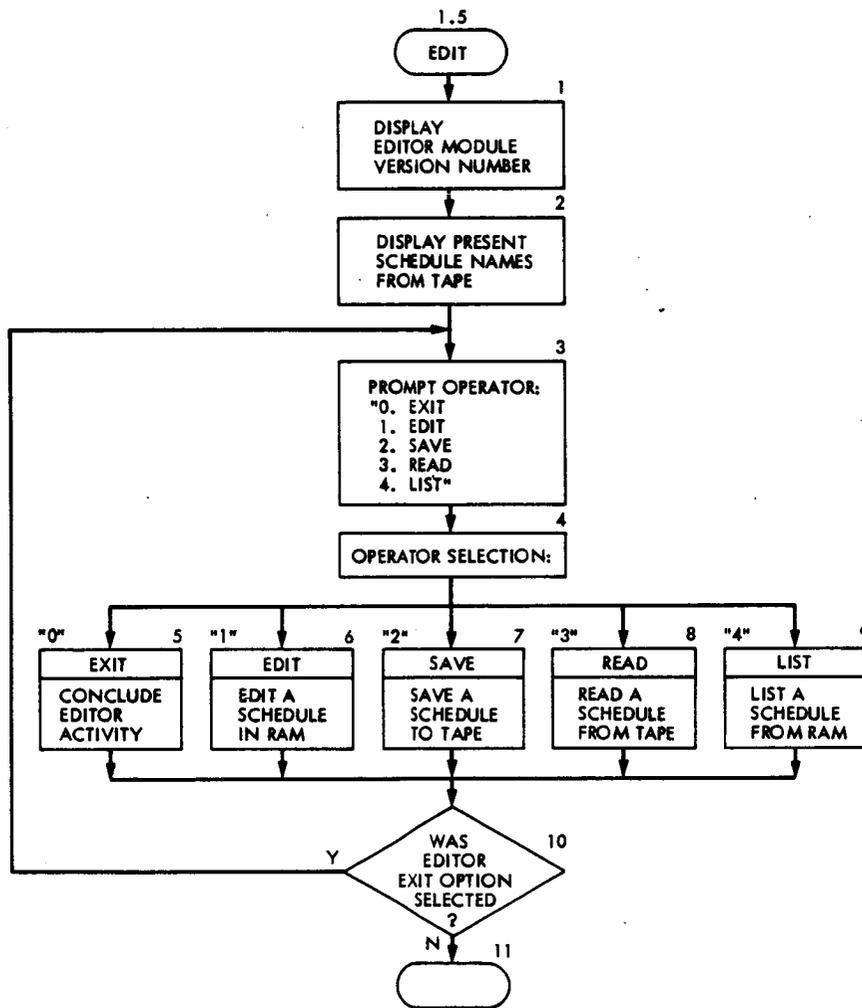


Fig. 11. EDITOR mode flowchart

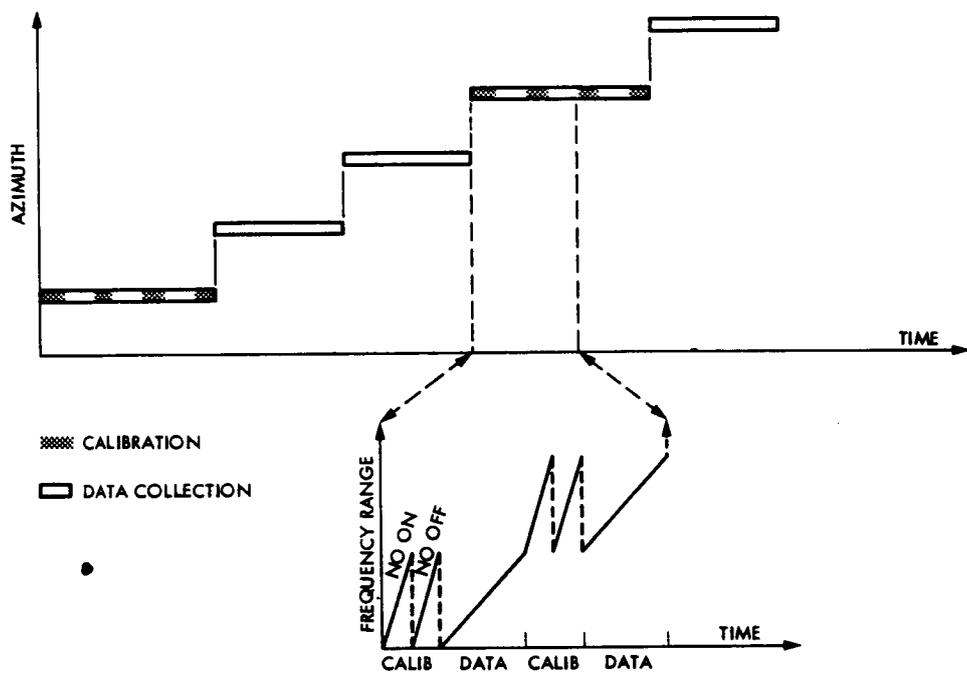


Fig. 12. Time history of system calibration as a function of azimuth

Mobile VLBI and GPS Measurement of Vertical Crustal Motion

P. M. Kroger and J. M. Davidson
Tracking Systems and Applications Section

E. C. Gardner
Kalamazoo College, Michigan

Mobile VLBI and GPS geodetic measurements have many error sources in common. Calibration of the effects of water vapor on signal transmission through the atmosphere, however, remains the primary limitation to the accuracy of vertical crustal motion measurements made by either technique. The two primary methods of water vapor calibration currently in use for mobile VLBI baseline measurements have been evaluated: radiometric measurements of the sky brightness near the 22 GHz emission line of free water molecules and surface meteorological measurements used as input to an atmospheric model. Based upon a limited set of 9 baselines, it is shown that calibrating VLBI data with water vapor radiometer measurements provides a significantly better fit to the theoretical delay model than calibrating the same data with surface meteorological measurements. The effect of estimating a systematic error in the surface meteorological calibration is shown to improve the consistency of the vertical baseline components obtained by the two calibration methods. A detailed error model for the vertical baseline component indicates current mobile VLBI technology should allow accuracies of order 5 cm with WVR calibration and 10 cm when surface meteorological calibration is used. A statistical analysis of the results of repeated measurements of the 336-km baseline from Big Pine, CA, to Pasadena, CA, shows the scatter in the vertical baseline component to be consistent with this model. A similar model for a hypothetical GPS baseline measurement is presented. A covariance analysis based upon this model shows current GPS technology to be capable of accuracies in the vertical baseline component comparable to present mobile VLBI measurements. Expected improvements in system components for both technologies should allow reduction of the uncertainty in the vertical component to less than 3 cm by 1989.

I. Introduction

The measurement of vertical crustal motion with microwave techniques such as Very Long Baseline Interferometry (VLBI) or ranging to Global Positioning System (GPS) satellites presents difficulties not encountered in the measurement

of horizontal motion. Inherent in the measurement of the vertical baseline component is a geometric dilution of precision (GDOP) arising from the fact that signals from the sources (quasars or satellites) can only be viewed along positive values of this coordinate whereas both positive and negative views of the sources are possible in measurements of horizontal base-

line components. In addition to this inherent geometrical weakness, the vertical baseline component is much more sensitive to errors in the calibration of tropospheric effects, particularly that portion due to atmospheric water vapor. This is especially true for the regional baselines measured by the mobile VLBI units. The relatively short length of these baselines causes essentially all of the tropospheric errors to map into the baseline vertical component. In the case of longer intercontinental baselines, a greater fraction of the tropospheric error would map into the horizontal components. To a large degree then, the precision which can be achieved in the measurement of vertical crustal motion rests upon how accurately we can calibrate the effects of the troposphere on the microwave signals.

Two techniques are currently in use for the calibration of the wet troposphere: a direct measurement of the 22-GHz emission of free water molecules along the antenna line of sight with a microwave radiometer and the use of an atmospheric model whose parameters are determined from surface meteorological measurements to infer the amount of water vapor along the line of sight. These two calibration methods have been compared using data from several recent mobile VLBI baseline measurements. In all cases it was found that calibrating data with WVR measurements provided a better fit to the theoretical delay model than did calibrating the same data with surface meteorological based estimates of atmospheric water vapor. As expected, the major differences in the values of the estimated baseline parameters for the two different calibrations were in the vertical baseline component where differences as large as 20 cm were seen.

The major sources of error in GPS-based vertical measurements are identical to those encountered in VLBI measurements. An important error source not found in VLBI measurements, however, is the uncertainty of the GPS satellite orbits. Described here is how data from a fiducial network of GPS receivers located at *a priori* well-known sites can be used to allow accurate determination of these orbits. Finally, the results of a covariance analysis of a simulated GPS baseline measurement in the Caribbean region are presented. These results show that the uncertainty in the wet troposphere calibration is the dominant source of error in the baseline vertical component just as in the case of mobile VLBI measurements.

II. Mobile VLBI Measurement of Vertical Motion

The essential components of a mobile or any other VLBI system are a pair of radio telescopes, high-precision frequency standards (hydrogen masers) at each antenna, and a special

purpose computer known as a correlator. During a baseline measurement, the broadband emission from an extragalactic radio source is recorded on some magnetic medium at each of the stations. The tapes from each station are then brought to the correlator which computes a cross correlation of the quasar signals recorded at the two stations. The resulting interferometric fringe pattern allows the difference in arrival times of the quasar signals at the two stations (delay) and its rate of change (delay rate) to be calculated. These delays and delay rates obtained from many separate quasar observations over a 24–36 hour period are used as input to a multiparameter least-squares estimation program that fits a detailed earth model to the observables from which the baseline coordinates and other parameters are extracted. More complete discussions of the interferometry technique can be found in Refs. 1–4 and of mobile VLBI measurements in Ref. 5.

III. Sources of Error in Mobile VLBI Vertical Measurements

Any deficiencies in either the earth model or the methods of calibration will ultimately degrade the accuracy of the estimated baseline components. Because baseline results are expressed in an earth-fixed system, the earth model must include effects due to the orientation and motion of the earth with respect to the coordinate frame of the quasars. These include the rate of earth rotation (UT1), precession, nutation, and polar motion. As has already been mentioned, the vertical component is particularly sensitive to the effects of propagation media on the delay observable. While the introduction of dual-frequency recording systems has essentially removed the effect of the ionosphere from mobile VLBI measurements, the calibration of the tropospheric delays, particularly that due to atmospheric water vapor, remains the single most important source of error in the measurement of the vertical baseline component. This is especially true for the relatively short regional baselines where tropospheric delays map almost entirely into this component. In addition to earth orientation and water vapor calibration, the measurement of the vertical baseline component is also subject to errors from several other sources including: instrumental phase calibration, receiver noise, and errors in the survey tie from the ground monument to the mobile antenna. Table 1 summarizes the error budget for the vertical baseline coordinate as it presently stands and how it may be expected to improve within the next few years. The error budget for the length coordinate is included for comparison.

A. Earth Orientation

Because we have detailed theoretical models which describe the orientation of the earth in space, it is, in principle, possible

to use the delay and delay rate observables to estimate these parameters along with the baseline parameters as part of the overall least-squares estimation process. Because the mobile VLBI baselines are relatively short, however, they would not provide as accurate a determination of earth orientation parameters as is presently available from external sources using results obtained with much longer baselines (Ref. 6). Presently earth orientation data from Ref. 7 are used, which are derived from a combination of Project POLARIS VLBI results (Ref. 8), VLBI measurements made by the antennas at the Deep Space Network stations (Ref. 9), and the TEMPO project at JPL (Ref. 10), and data from lunar laser ranging experiments (Ref. 11). Data from these sources are combined with a Kalman filter based upon studies of atmospheric effects on the earth's angular momentum budget. The values of UT1 determined from this analysis are input explicitly to the parameter estimation program. Present accuracies of these UT1 values lie in the range of 0.1 to 0.6 milliseconds. In the case of polar motion data, presently obtained from BIH Circular D (Ref. 24), accuracies of 0.007 arc seconds were assumed for both the x and y components. The resulting uncertainties in the estimated vertical coordinate from these two sources are shown in Table 1 for a typical regional baseline of 300 to 500 km in length.

Expected improvements in both polar motion and earth rotation data should reduce the contributions of these sources to the levels shown in the second column of Table 1 by 1989. These improvements will result from the upgrades to the VLBI systems of the POLARIS project, addition of new lunar laser ranging stations, and incorporation of earth orientation data obtained from laser ranging to the Lageos satellite.

B. Troposphere Calibration

It is evident from Table 1 that the tropospheric portion of the delay model represents the most serious concern for estimation of vertical baseline components with mobile VLBI measurements. Although it is possible in principle to estimate the tropospheric delay (using the mobile VLBI data) as part of the overall parameter estimation process, it is undesirable for several reasons. First, there may be large spatial and temporal inhomogeneities in the troposphere which are virtually impossible to model realistically. Second, for the relatively short regional baselines monitored in the Mobile VLBI Program, the troposphere and baseline vertical parameters are strongly correlated, resulting in greatly reduced precision in the estimation of the latter. Third, correlations notwithstanding, the requirement that the data be used to estimate the tropospheric delay will seriously weaken its strength for baseline estimation. For these reasons, the effect of the troposphere on the quasar signals must be removed prior to parameter estimation by an external means of calibration.

The effect of the dry component of the troposphere (i.e., all nitrogen and oxygen) on the VLBI data is calibrated using a combination of surface meteorological (SM) measurements used as input to an atmospheric model (Ref. 12). Errors in this calibration approach arise principally from the facts that the atmosphere is not in hydrostatic equilibrium, that there may be large horizontal gradients in barometric pressure, and that the line-of-sight delay must be inferred from the zenith delay using a model-dependent mapping function (Ref. 13). Under extreme conditions, these could combine to contribute a systematic error of up to 3 cm in the baseline vertical, although more typically the contribution is of order 1 cm (Ref. 5).

The calibration of the wet troposphere poses a more serious problem. Surface humidity measurements may bear little relation to the distribution of water vapor at altitude. Furthermore, since azimuthal symmetry in the distribution of the water vapor cannot be assumed, accurate mapping to the antenna line of sight cannot be guaranteed. For these reasons considerable effort and resources have gone into development of water vapor radiometers (WVRs) to directly measure the quantity of water vapor along the antenna line of sight (Refs. 14-16).

The WVRs currently in use for mobile VLBI measurements, fabricated by the NASA Crustal Dynamics Project (CDP) (Ref. 17), are the result of this ongoing effort. These devices measure the intensity of microwave emission near 22 GHz from free water molecules along the antenna line of sight. This data is then used to calculate an equivalent zenith path delay with an accuracy of 1 to 2 cm. Because the radiometer can be pointed with the telescope, the effects of spatial and temporal nonuniformity are removed. Future developments in WVR technology (Ref. 18) are expected to reduce the contribution of the wet troposphere to the vertical baseline uncertainty to less than 1 cm.

To illustrate the beneficial effect that use of water vapor radiometer calibration can have on VLBI measurements, nine baselines were compared where both surface meteorological and water vapor radiometer data were available. The mobile VLBI data were recorded during the CDP observing sessions of June 6, June 29, and August 28 of 1983. The Chao model (Ref. 12) was used to calibrate path delays using surface meteorological data, and the method of Resch (Ref. 25) was used to calibrate path delays from the WVR data. The VLBI data were then processed twice, using each of these calibrations to produce independent solutions. Table 2 contains a comparison of the rms scatters of the post-fit residuals resulting from the least squares fit to the delay model from these two sets of analyses. As this table shows, the use of the WVR data results in a better fit to the delay model (and hence a

lower rms scatter) for all 9 baselines in the comparison. A simple analysis of the improvement indicates that for 6 of the 9 baselines, the use of WVR data represents a statistically significant improvement over surface meteorology data as a means of wet troposphere calibration. It should be noted that this analysis assumes no correlation between the rms scatters for the two calibration methods. Since, in fact, they are highly correlated (due to the fact that they have all but one error source in common), these levels of significance represent lower limits.

The effect of wet troposphere calibration on the actual values of the estimated baseline parameters is shown in Fig. 1. As can be seen, the largest effect is in the vertical baseline component where differences of up to 20 cm are seen between the SM and WVR approaches. As expected, the method of calibration is shown to have a significant effect on the estimated values of this component. In the horizontal components, the impact of changing the calibration method is much less severe.

Finally, we note that for longer baselines, the correlation of the troposphere and the baseline vertical is considerably reduced and estimation of an error in the zenith path delay may improve the results by accounting for any systematic error in the troposphere calibration. This may be especially true when surface meteorology calibration of the wet troposphere has been used. In this work, station zenith errors were estimated for several baselines from the set of 9 used in the above comparison (see Table 2), and it was found that the estimation of a constant zenith error for results calibrated with surface meteorology data does produce a significant improvement in the consistency of the estimated baseline parameters. This is especially true for the case of the baseline vertical component. Figure 2 shows the effect that the estimation of this additional parameter has on the length and vertical components for six baselines from our original comparison set. In all cases the estimated baseline parameters tend to move toward the results obtained with WVR calibrated data indicating that the average difference between the surface meteorology and WVR-based results is significantly reduced when this additional parameter is estimated. If the WVR-based solution is assumed to represent the "true" baseline, then this estimation procedure is seen to improve repeatability by a factor of over 2 in both the length and vertical components. An unfortunate (but expected) consequence of estimating a zenith troposphere error is that the formal uncertainty in the vertical baseline component is increased by a factor of between 2 and 3. This is a result of a still high correlation between the baseline vertical parameter and the estimated troposphere parameter and a result of the simple fact that the same amount of data is now required to estimate a larger number of parameters. Figure 3 shows how the formal uncertainties in the

length and vertical parameters increase when this constant zenith offset parameter is estimated.

The calibration techniques described in this subsection are all suboptimal in various senses. The WVR-based calibration is clearly superior to the SM-based calibration. But the direct application of WVR calibration makes no allowance for the effect of WVR errors on the VLBI baseline results. The alternative approach of troposphere estimation improves the consistency of the VLBI results, but it also results in a serious loss of precision. A third alternative which combines the strengths of both the above is the hybrid approach in which WVR calibration is applied to the raw VLBI observables and a constrained calibration error is estimated, based on a realistic model of that error. The constraint is applied in the form of an *a priori* covariance, which limits the size of the error model parameters, in accordance with the demonstrated inherent accuracy of the WVR instrument. This hybrid approach is under development at JPL (Ref. 19) and testing with data is expected to begin soon.

C. Computation of Uncertainties

In lieu of an operational hybrid approach to computation of uncertainties, as described in the preceding subsection, it has not been a practice to attempt to estimate any sort of systematic errors in the troposphere calibration of the mobile VLBI data. However, since all calibration is subject to systematic error, we add additional uncertainty to the post-fit estimated baseline parameters in order to provide more realistic values for their uncertainties. In the cases where water vapor radiometer data is available, we add 2.0 cm in quadrature to the uncertainty in the local vertical coordinate for each station. To reflect the larger error expected when surface meteorological data is used for wet troposphere calibration, an uncertainty equal to one half the mean wet path delay is added to the local vertical for that station. Typically this results in an additional error of from 2.5 to 5.0 cm in the local vertical coordinate. The resulting uncertainties in the vertical baseline coordinate for a typical regional baseline are shown in Table 1. Expected improvements in water vapor technology should allow reduction of the contribution of the wet troposphere to less than 2.0 cm provided a sufficient number of these new instruments are made available for use at all observing stations.

Table 1 also contains the contributions from several other error sources which have some effect on the accuracy of the vertical baseline component including an estimate of the effects of unmodeled elements of the delay observable such as ocean loading and antenna flexure. The combined effects of these additional error sources, however, represent only a small fraction of the contributions from earth orientation and wet troposphere calibration to the uncertainty in the vertical

component. A more extensive discussion of these and other sources of error in mobile VLBI measurements can be found in Ref. 5.

In order to test the validity of this error model for the vertical baseline component, the results of repeated measurements of a single baseline have been statistically analyzed. The 336-km baseline from the 40-meter telescope at the Owens Valley Radio Observatory located near Big Pine, CA to the mobile VLBI site at the Jet Propulsion Laboratory in Pasadena, CA has been measured 19 times since the first measurement in January 1980. From these 19 measurements we have taken a subset of 15 measurements consisting of 4 groups, each of which contains measurements made within a six month period. This was done to minimize the possible effects of tectonic motion on these tests of consistency for this baseline which crosses the San Andreas fault. The results of these measurements are shown in Fig. 4 in the form of error ellipses for baseline length and vertical components. All results are expressed as shifts from a nominal reference baseline. It is clear from the figure that the repeatability in the vertical coordinate is much worse than in the baseline length. This is not unexpected since the baseline length component is entirely unaffected by uncertainties in the earth orientation and only slightly affected by troposphere calibration errors. It should also be noted that some of these baselines used surface meteorological data to calibrate the wet troposphere and that in these cases no attempt was made to estimate a systematic error in the manner described earlier. This, as well as inadequacies in the current dry troposphere model, may be responsible for the poor repeatability in some cases.

From a χ^2 analysis of these results, shown in Table 3, it is evident that the error model for the vertical baseline component is adequate for all but the second group of measurements where the scatter in the measurements is somewhat larger than would be expected on the basis of the formal errors. This is surprising, because all measurements in this group used water vapor radiometers to calibrate the wet troposphere delay. This may indicate that there exists a substantial systematic error in some of the WVR measurements which might be removed by estimating a zenith path delay error as was done for the surface meteorological data with some success. Beckman (Ref. 19) has proposed such a model, which is currently being evaluated at JPL.

The results contained in Table 3 also serve to illustrate the general improvement in consistency of repeated measurements since 1980. This is due to many factors including improvements in receiver technology, theoretical models, data processing software, as well as improved wet troposphere calibration. A more extensive discussion of these improvements is contained in Ref. 5.

IV. GPS Measurement of Vertical Crustal Motion

A. Measuring Baselines with GPS Receivers

By 1989 the full 18 satellite constellation of GPS/NAVSTAR satellites is expected to be in orbit (Ref. 20). These satellites will broadcast phase modulated signals at two L-band frequencies (1575.41 MHz and 1227.60 MHz) and are expected to have many civilian users for navigation, precise orbiter tracking and geodesy. Among the applications being investigated at the Jet Propulsion Laboratory is a GPS based system to measure geodetic baselines with accuracies comparable to those attainable using present mobile VLBI systems.

The basic observable obtained from all GPS measurements is the distance from the receiver to the satellite, called the "pseudorange." It is not exactly equal to the true satellite-receiver distance because of offsets between the satellite and receiver clocks and the effects of the troposphere and ionosphere on the broadcast signals. The covariance analysis described here, however, does not use the pseudorange directly but forms a new datum by taking the difference between the pseudoranges from two receivers to a single GPS satellite. Use of this differential pseudorange has the advantage of removing the clock offset term between the receiver and satellite. This still leaves an offset between the two receiver clocks, but it is also possible to form a second difference between two differential pseudoranges to two different GPS satellites. This second difference has the effect of canceling out the clock errors between the two ground receivers, thus removing all errors due to clock offsets from the GPS range data. When there are more than two receivers involved, a more elaborate linear combination of the pseudoranges is actually used to remove the clock offsets (Ref. 21). It is this more general double differencing technique which has been used in the covariance analysis presented here.

B. Fiducial Network Concept

Although the double differencing process removes all clock errors from the data, the uncertainty in the satellite ephemeris would limit accuracy to a decimeter or greater for baselines longer than 100 km. For this reason the concept of a fiducial network has been developed (J. L. Fanelow and J. B. Thomas, personal communication, 1983 and Ref. 22). The fiducial network consists of three or more GPS receivers whose locations have been accurately established by an independent technique such as VLBI or satellite laser ranging. During a GPS baseline measurement, receivers at *a priori* unknown locations observe the satellites simultaneously with the receivers located at the precisely known fiducial station locations. In the subsequent least squares analysis, the range data from the fiducial station receivers allows accurate values of

the GPS orbits to be estimated along with values for the unknown station locations. Without this ability to estimate accurate GPS orbits, centimeter-level baseline accuracies for regional baselines longer than 100 km in length would not be possible. The schematic diagram in Fig. 5 shows a GPS measurement scenario using a 4-station fiducial network in conjunction with 2 mobile GPS receivers. Initially, fiducial sites will be located at Ft. Davis, TX, Westford, MA, and Richmond, FL. Other fiducial sites being considered are Quito, Ecuador; Sao Paulo, Brazil; Santiago, Chile and Cayenne, French Guiana (Ref. 23).

V. Sources of Error in GPS Vertical Measurements

The primary sources of error in VLBI vertical measurements described earlier are also present in GPS-based measurements of vertical crustal motion. GPS measurements have the additional error source of uncertainties in the GPS orbits. Use of a fiducial network of observing stations in the manner described in Ref. 23 eliminates most of the error resulting from this source for baselines of less than 1000 km in length. This, of course, requires that we have adequate models of the satellite orbits which include effects of solar radiation pressure, higher order gravitational terms, uncertainty in the value of the gravitational constant and that we properly account for any relativistic effects. Table 4 contains the major elements of the error model assumed in our covariance analysis.

Table 5 shows the results of this analysis for the 370-km baseline between San Juan, Puerto Rico and Santo Domingo, Dominican Republic for both the current and the 1989 system components. As was the case for mobile VLBI vertical measurements, it is the troposphere calibration which dominates the uncertainty in the vertical baseline component. The approximately 8-cm uncertainty for the current GPS system is, in fact, comparable to the accuracy now achieved with mobile VLBI systems. It is clear from these results that improvements in the accuracy of the wet troposphere calibration are essential if vertical component accuracies of 2 to 3 cm are to be

achieved with GPS technology. It is encouraging to note that errors associated with the determination of GPS orbits do not contribute significantly to baseline errors for baselines of this length.

VI. Conclusions

Four years of baseline measurements with mobile VLBI systems have shown that baseline vertical components can be measured with a precision on the order of 5 cm. The major sources of error in these measurements are uncertainty in the earth orientation, especially in the position of the pole, and errors in the calibration of the effects of water vapor in the troposphere. The wet troposphere calibration presently represents the most serious error source for mobile VLBI measurements of vertical crustal motion. Future improvements in WVRs promise to reduce the contribution of water vapor radiometer calibration errors to a level of 2 cm or less. Along with improvements in the knowledge of the earth's polar motion, this will allow baseline vertical components to be measured with a precision of 2 to 3 cm. The introduction of a WVR error model into the parameter estimation process may allow the effects of systematic errors in the wet troposphere calibration to be reduced even further, to the point that the calibration of the dry troposphere, including the effects of dynamic terms and horizontal gradients, may become the limitation on accuracy.

Baseline measurements using GPS satellites and receivers have the potential of achieving VLBI level accuracies in both the horizontal and vertical coordinates. The measurement of vertical motion with GPS technology will also require the use of water vapor radiometers to remove the effects of the wet troposphere from the satellite/receiver range measurements. Present accuracies in the vertical coordinate are predicted, on the basis of a covariance analysis, to be comparable to the present accuracies of mobile VLBI measurements. Improvements in receiver technology, WVR instrumentation, and GPS orbit determination could conceivably reduce the uncertainty in vertical measurements to the order of 2 to 3 cm by 1989.

References

1. Thomas, J. B., An Analysis of Long Baseline Radio Interferometry, *JPL Technical Report 32-1526*, vol. 7, 37-50; Jet Propul. Lab., Pasadena, Calif., 1972a.
2. Thomas, J. B., An Analysis of Long Baseline Radio Interferometry, II, *JPL Technical Report 32-1526*, vol. 8, 29-38; Jet Propul. Lab., Pasadena, Calif., 1972b.
3. Thomas, J. B., An Analysis of Long Baseline Radio Interferometry, *JPL Technical Report 32-1526*, vol. 16, 47-64; Jet Propul. Lab., Pasadena, Calif., 1974.
4. Shapiro, I. I., Estimation of Astrometric and Geodetic Parameters from VLBI Observations, in *Methods of Experimental Physics*, M. L. Meeks, ed., 261-276, Academic Press, New York, 1976.
5. Davidson, J. M., and D. W. Trask, Utilization of Mobile VLBI for Geodetic Measurements, *IEEE Transactions on Remote Sensing and Geoscience, Special Issue on Geodynamics*, 1985 (in press).
6. Allen, S. L., Earth Orientation Effects on Mobile VLBI Baselines, *TDA Progress Report 42-78*, pp. 202-206, Jet Propul. Lab., Pasadena, Calif., 1984.
7. Eubanks, T. M., J. A. Steppe, M. A. Spieth, The Accuracy of Radio Interferometric Measurements of Earth Rotation, *TDA Prog. Rept. 42-80*, 229-235, Jet Propul. Lab., Pasadena, Calif., 1984.
8. Carter, W. E., D. S. Robertson, J. E. Pettey, B. D. Tapley, B. E. Schutz, R. J. Eanes, and M. Lufeng, Variations in the Rotation of the Earth, *Science*, 224, 957-961, 1984.
9. Sovers, O. J., J. B. Thomas, J. L. Fanelow, E. J. Cohen, G. H. Purcell, Jr., D. H. Rogstad, L. J. Skjerve, and D. J. Spitzmesser, Radio Interferometric Determination of Intercontinental Baselines and Earth Orientation Utilizing Deep Space Network Antennas, *J. Geophys. Res.* 89, 7597-7607, 1984.
10. Eubanks, T. M., M. G. Roth, P. B. Esposito, J. A. Steppe, and P. S. Callahan, An Analysis of Tempo Earth Orientation Results, *Proceedings of Symposium No. 5: Geodetic Applications of Radio Astronomy*, edited by William Carter, U. S. Department of Commerce, National Oceanic and Atmospheric Administration, 152-162, 1982.
11. Dickey, J. O., and J. G. Williams, Earth Rotation from Lunar Laser Ranging, *Astron. Astrophys. Suppl. Ser.*, 54, 519-540, 1983.
12. Chao, C. C., A New Method to Predict Wet Zenith Range Correction from Surface Measurements, *DSN Progress Report 32-1526*, vol. 14, pp. 33-41, Jet Propul. Lab., Pasadena, Calif., 1973.
13. Lanyi, G. A., Tropospheric Delay Effects in Radio Interferometry, *TDA Prog. Rept. 42-78*, 152-159, Jet Propul. Lab., Pasadena, Calif., 1984.
14. Claflin, E. S., S. C. Wu, and G. M. Resch, Microwave Radiometer Measurement of Water Vapor Path Delay: Data Reduction Techniques, *DSN Progress Report 42-48*, pp. 22-30, Jet Propul. Lab., Pasadena, Calif., 1978.
15. Wu, S. C., Optimum Frequencies of a Passive Microwave Radiometer for Tropospheric Path Length Correction, *IEEE Transactions on Antennas and Propagation AP-27, No. 2*, 233-239, 1979.

16. Resch, G. M., M. C. Chavez, and N. I. Yamane, Description and Overview of an Instrument Designed to Measure Line-of-Sight Delay Due to Water Vapor, *TDA Prog. Rept. 42-72*, 1-19, Jet Propul. Lab., Pasadena, Calif., 1982.
17. NASA, Application of Space Technology to Crustal Dynamics and Earthquake Research, *NASA Technical Paper 1464*, NASA Office of Space and Terrestrial Applications, Washington, D. C., 1979.
18. Janssen, M. A., A New Instrument for the Determination of Radio Path Delay Variations Due to Atmospheric Water Vapor, *IEEE Transactions on Remote Sensing and Geoscience, Special Issue on Geodynamics*, 1985 (in press).
19. Beckman, B. C., A WVR Error Model, *IEEE Transactions on Geoscience and Remote Sensing, Special Issue on Geodynamics*, 1985 (in press).
20. Brady, W. F. and P. S. Jorgensen, Worldwide Coverage of the Phase II NAVSTAR Satellite Constellation, *Navigation 28(3)*, 167-177, 1981.
21. Wu, J. T., Elimination of Clock Errors in a GPS Based Tracking System, Paper presented at the AIAA/AAS Astrodynamics Conf., AIAA-84-2052, Seattle, Washington, August 20, 1984.
22. Davidson, J. M., C. L. Thornton, B. C. Beckman, P. M. Kroger, S. A. Stephens, and S. C. Wu, Covariance Analysis for Geodetic Measurements Utilizing the Global Positioning System Satellites (in preparation), 1985.
23. Kroger, P. M., C. L. Thornton, J. M. Davidson, S. A. Stephens, and B. C. Beckman, Sensitivity of GPS Caribbean Baseline Performance to the Location of a Southerly Fiducial Station, *Proceedings of the First International Symposium on Precise Positioning with the Global Positioning System*, Rockville, MD, April 15, 1985 (in press).
24. Feissel, M., *Annual Report for 1982*, Paris, BIH (Bureau International l'Huere), 1983.
25. Resch, G. M., Inversion Algorithms for Water Vapor Radiometers Operating at 20.7 and 31.4 GHz, *TDA Progress Report 42-76*, Jet. Propul. Lab., Pasadena, Calif., 1983.

Table 1. Summary of mobile VLBI error sources in length and vertical coordinates^a

| Error Source | Contribution to Uncertainty
in Baseline Length and Vertical Coordinate, cm | | | |
|--|---|------------------|------------------|------------------|
| | 1984 System | | 1989 System | |
| | Length | Vertical | Length | Vertical |
| Earth Orientation | | | | |
| UT1 - UTC | 0 | 0.5 | 0 | 0.2 |
| Polar Motion | 0 | 1.5 | 0 | 1.0 |
| Propagation Media | | | | |
| Dry Troposphere | 0.8 | 1.5 | 0.5 | 1.0 |
| Wet Troposphere ^b | 0.6 | 4.0 | 0.3 | 2.0 |
| Wet Troposphere ^c | 1.2 | 9.0 | 1.2 | 9.0 |
| Miscellaneous | | | | |
| Source Positions | 0.3 | 0.3 | 0.3 | 0.3 |
| System Noise | 0.2 | 0.2 | 0.2 | 0.2 |
| Mobile Antenna Survey Tie
to Monument | 0.3 | 0.5 | 0.3 | 0.5 |
| Various Unmodeled
Elements of Delay Model | 0.5 | 1.0 | 0.2 | 0.5 |
| Total (RSS) | 1.2 ^b | 4.7 ^b | 0.8 ^b | 2.6 ^b |
| | 1.7 ^c | 9.3 ^c | 1.4 ^c | 9.1 ^c |

^aThe contributions to uncertainties in the length and vertical baseline components shown in this table are meant to represent those expected for a typical regional baseline of 300 to 500 km in length measured with one mobile and one fixed antenna. Actual values will depend upon baseline length and orientation, and the performance of all system components.

^bWater vapor radiometer data used to calculate wet path delays.

^cSurface meteorological data and atmospheric model used to calculate wet path delays.

Table 2. Values of post-fit residual delay scatter

| Experiment Date (UT) | Stations ^a | Baseline Length, km | RMS Scatter, ps | | Probability That RMS Are the Same, % ^b |
|----------------------|-----------------------|---------------------|-----------------|--------------|---|
| | | | WVR Calibration | Surface Met. | |
| 06/09/83 | HAT CREEK GRAS | 1933.5 | 87.3 | 139.2 | 0.04 |
| | GRAS WESTFORD | 3134.9 | 115.6 | 167.0 | 0.10 |
| | HAT CREEK WESTFORD | 4032.8 | 124.3 | 167.8 | 1.20 |
| 06/29/83 | OVRO 130 HAT CREEK | 484.3 | 48.8 | 104.6 | 4.0×10^{-6} |
| | OVRO 130 MOJAVE | 245.3 | 61.0 | 163.5 | 1.0×10^{-5} |
| | HAT CREEK MOJAVE | 729.1 | 42.2 | 106.7 | 4.0×10^{-7} |
| 08/28/83 | OVRO 130 MOJAVE | 245.3 | 104.0 | 120.5 | 36.0 |
| | MOJAVE DSS 13 | 12.6 | 197.9 | 204.2 | 84.0 |
| | OVRO 130 DSS 13 | 257.6 | 151.4 | 159.1 | 68.0 |
| | | | | | |

- ^aDSS 13 - Venus Antenna, NASA Deep Space Network Complex, Goldstone, CA
 GRAS - George R. Agassiz Radio Astronomy Station, Ft. Davis, TX
 HAT CREEK - Hat Creek Radio Astronomy Observatory, Cassel, CA
 MOJAVE - Mojave Base Station, NASA Deep Space Network Complex, Goldstone, CA
 OVRO 130 - Owens Valley Radio Observatory, Bishop, CA
 WESTFORD - Haystack Observatory, Westford, MA

^bCalculated assuming that residual scatter is Gaussian and that the values from the two different calibration methods are independent. Because this is not true, these numbers represent an upper limit.

Table 3. Owens Valley → JPL baseline precision^a

| Dates | Degrees of Freedom | RMS Scatter, cm | | χ^2 | |
|--------------|--------------------|-----------------|----------|----------|----------|
| | | Length | Vertical | Length | Vertical |
| 5/80 - 11/80 | 3 | 2.2 | 10.2 | 0.33 | 0.55 |
| 8/81 - 11/81 | 3 | 0.8 | 12.6 | 0.28 | 2.29 |
| 10/82 - 2/83 | 3 | 1.1 | 5.9 | 0.83 | 0.17 |
| 6/83 - 11/83 | 2 | 0.8 | 5.6 | 0.40 | 0.17 |

^aThe data from which these results are calculated are displayed in Fig. 4.

Table 4. Inputs to GPS baseline covariance analysis

| Input | 1984 System | 1989 System |
|---------------------------------------|---|---|
| Data Type | Double differenced integrated Doppler | Double differenced range or integrated Doppler |
| Data Noise | 3.0 cm | 1.0 cm |
| Fiducial Network | Ft. Davis, TX
Westford, MA
Richmond, FL | Ft. Davis, TX
Westford, MA
Quito, Ecuador |
| <u>Considered Parameters</u> | | <u>Uncertainty</u> |
| Relative Position of Fiducial Station | 3 cm horizontal
9 cm vertical | 1 cm all components |
| Troposphere Calibration Error | 2.0 cm (zenith) | 0.75 cm (zenith) |
| Solar Radiation Pressure | | 5% @ $C_R = 0.5$ |
| Geopotential | | 10% of GEM 6 - APL50 |
| GM | | 10^{-8} of the total value |
| <u>Estimated Parameters</u> | | <u>A Priori Uncertainty</u> |
| Range Bias Parameters | | 1000 m (integrated Doppler only) |
| NAVSTAR Position | | 10 m in each component (X, Y, Z) |
| NAVSTAR Velocity | | 0.1 cm/s in each component (V_x, V_y, V_z) |
| Mobile Station Location | | 20.0 m in each component (X, Y, Z) |

Table 5. Covariance analysis results for San Juan to Santo Domingo baseline^a

| Error Source | Contribution to Uncertainty
Baseline Vertical Coordinate, cm | |
|---|---|--------------------------|
| | 1984 System | 1989 System ^b |
| Range Measurement Accuracy | 3.4 | 1.4 |
| Solar Radiation Pressure | << 1 | << 1 |
| Geopotential | << 1 | << 1 |
| GM | << 1 | << 1 |
| Relative Position
of Fiducial Stations | 0.8 | 0.1 |
| Location of Geocenter | << 1 | << 1 |
| Troposphere Calibration | 6.9 | 2.6 |
| Total (RSS) | 7.7 | 2.9 |

^aThe values in this table are based upon the results for a single baseline of 370 km between San Juan, Puerto Rico and Santo Domingo, Dominican Rep. The values for other baselines will depend to a degree upon baseline length and orientation.

^bThese results correspond to the integrated Doppler data type. Results for the carrier range data type are virtually identical for this short baseline.

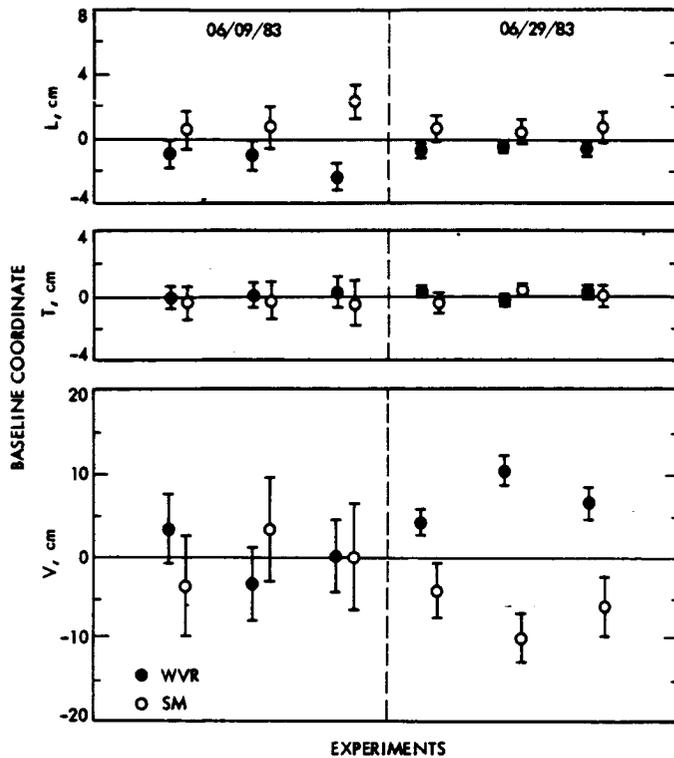


Fig. 1. The effect of different wet troposphere calibrations on estimated baseline parameters is illustrated for 6 baselines from two mobile VLBI experiments completed in June 1983. Estimated values of the length, transverse, and vertical baseline components for both WVR and surface meteorological calibrations are shown. See Table 2 for station names and baseline lengths.

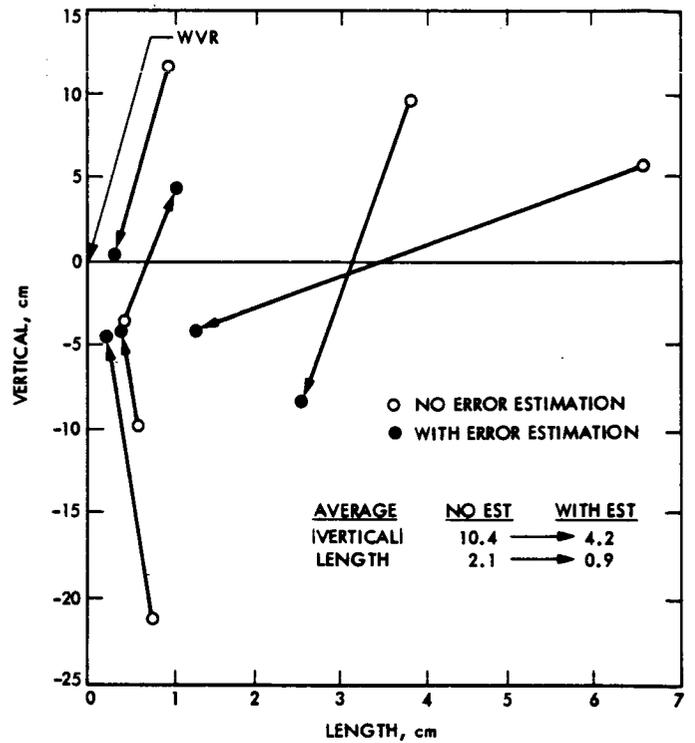


Fig. 2. The effect of the estimation of zenith troposphere error on length and vertical coordinates is shown. The difference between the WVR and SM values is substantially reduced by estimating an error in the SM calibration.

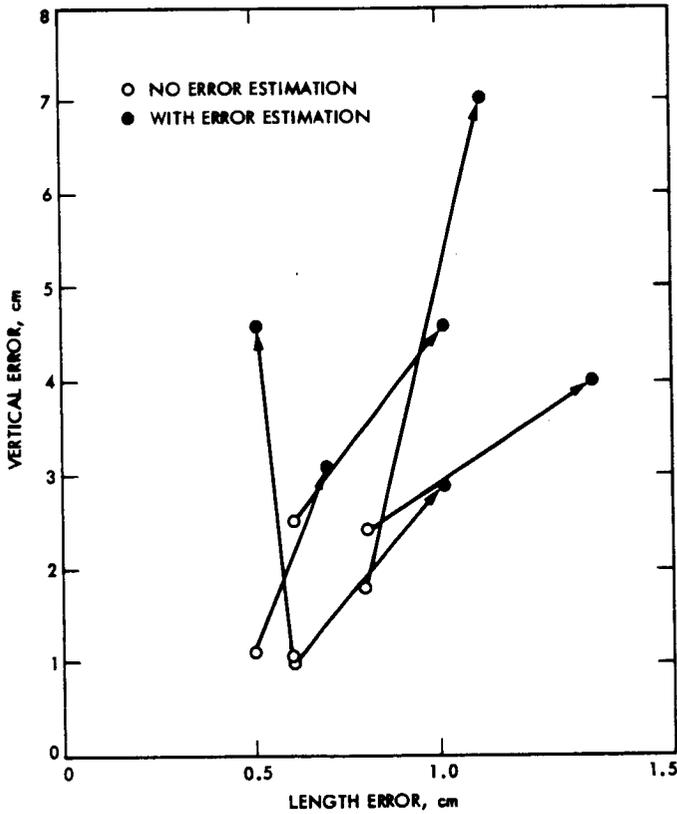


Fig. 3. The effect of the estimation of zenith troposphere error on formal uncertainties in the length and vertical coordinates is shown. The accuracy of these components is substantially increased by the estimation of an error in the calibration.

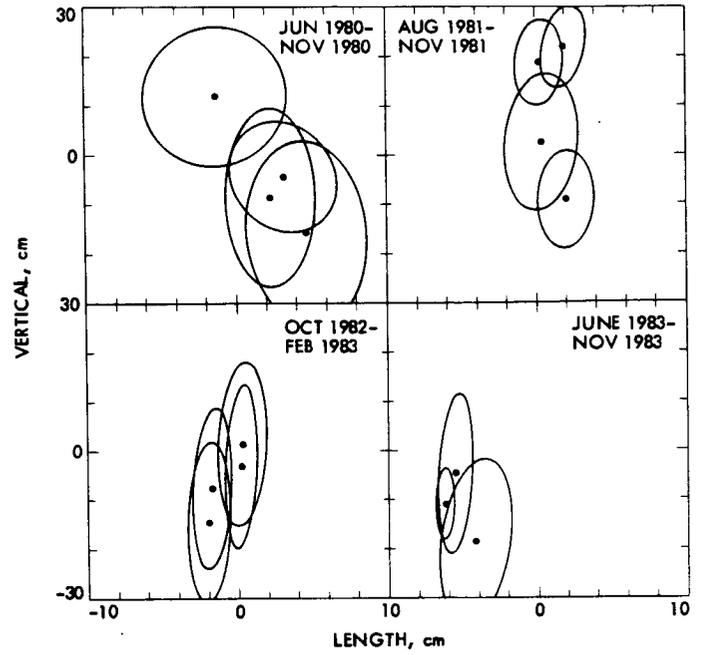


Fig. 4. Length and vertical coordinates of baselines from Big Pine, CA to Pasadena, CA are shown as shifts from a nominal reference baseline. Note how the rms scatters and the formal errors decrease with time (see Table 3).

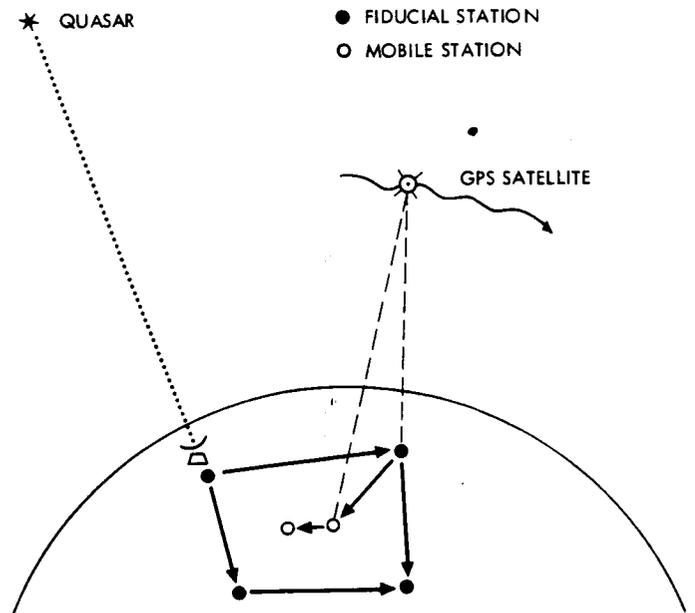


Fig. 5. The use of a fiducial network for GPS baseline measurements is shown. The fiducial station data enable the accurate estimation of the GPS satellite orbits. This, in turn, enables the accurate determination of the mobile station locations. VLBI observations establish the fiducial station baselines and tie the GPS results to the inertial frame of the quasars.

Improved Mapping of Radio Sources From VLBI Data by Least-Squares Fit

E. R. Rodemich

Communications Systems Research Section

This article describes a method for producing improved mapping of radio sources from VLBI data. The method described here is more direct than existing Fourier methods, is often more accurate, and runs at least as fast. The visibility data is modeled here, as in existing methods, as a function of the unknown brightness distribution and the unknown antenna gains and phases. We want to choose these unknowns so that the resulting function values are as near as possible to the observed values. If we use the RMS deviation to measure the closeness of this fit to the observed values, we are led to the problem of minimizing a certain function of all the unknown parameters. This minimization problem cannot be solved directly, but it can be attacked by iterative methods which we show converge automatically to the minimum with no user intervention. The resulting brightness distribution will furnish the best fit to the data among all brightness distributions of given resolution.

I. Introduction

This article describes a method for producing improved mapping of radio sources from VLBI data. The use of VLBI data has led to high resolution maps of radio sources in the sky [Refs. 1-4]. The data provide values of the visibility function, which is the Fourier transform of the brightness distribution. The problem of finding the unknown brightness distribution can accordingly be expressed as the problem of finding an inverse Fourier transform. The methods currently used depend on approximate inversion methods for a Fourier transform which is known on an irregularly spaced set of points.

An additional complication is that the signal received at each antenna can have an unknown gain and phase offset, depending on conditions at this antenna as well as on atmo-

spheric conditions. This introduces unknown multiplicative factors into the visibilities which must be eliminated before inverting the Fourier transform. Iterative methods have been developed for this which use an assumed map to recalibrate the data, get a new map from these data by Fourier inversion, then repeat the procedure starting from the new map. These iterative methods require considerable user interaction as well as computer time. They are also biased in favor of certain types of brightness distributions in the resulting map.

The method described in this article is more direct than existing Fourier methods, is often more accurate, and runs at least as fast. The visibility data are modeled here, as in existing methods, as a function of the unknown brightness distribution and the unknown antenna gains and phases. We want to choose these unknowns so that the resulting function values are as near as possible to the observed values. If we use the

RMS deviation to measure the closeness of this fit to the observed values, we are led to the problem of minimizing a certain function of all the unknown parameters. This minimization problem cannot be solved directly, but it can be attacked by iterative methods which converge automatically to the minimum with no user intervention. The resulting brightness distribution will furnish the best fit to the data among all brightness distributions of given resolution.

II. The Method

Preprocessing of the data, which we are not concerned with here, furnishes time averages of the visibility function over intervals on the order of several seconds to a minute, together with estimates of standard deviation due to noise. We get a set of values, $E_{pq}^{(n)}$, for the antenna pair p, q and the n^{th} time interval. For a given value of n , only some of the possible p, q pairs may occur, either because the source is not visible from all antennas, or because some data were lost.

Each data value, $E_{pq}^{(n)}$, is modeled by a function, $F_{pq}^{(n)}$, of the unknown brightness distribution, I_m , and the unknown gains, $A_p^{(n)}, A_q^{(n)}$, with noise added. The form of $F_{pq}^{(n)}$ is given in Appendix A. Using the standard deviations, $\sigma_{pq}^{(n)}$, from the preprocessing, we set up the function

$$Q = \sum_n Q_n,$$

$$Q_n = \sum_{p,q} \left| F_{pq}^{(n)} - E_{pq}^{(n)} \right|^2 / \sigma_{pq}^{(n)2}.$$

Each $F_{pq}^{(n)}$ is linear in the I_m 's and contains the product $A_p^{(n)} \overline{A_q^{(n)}}$, where the bar denotes a complex conjugate. An iterative method of minimizing Q is used which repeats step (1) followed by step (2):

- (1) Minimize Q by varying the I_m 's, holding the $A_p^{(n)}$'s fixed.
- (2) Minimize Q by varying the $A_p^{(n)}$'s, holding the I_m 's fixed.

The nature of these steps is quite different. In (1), we want to solve a system of simultaneous linear equations in a large number of unknowns. In (2), this linearity is lost, but the problem breaks up into the minimization of each Q_n separately, which only involves a small number of variables.

As this iterative procedure is carried out, the value of Q always decreases, and the unknowns approach values for which no further decrease is possible. Hence the procedure must converge.

III. Results and Conclusions

The method was applied to two sets of real data and one set of simulated data. The two real data sets are based on observations in Dec. 1982 at 6 cm. The resulting maps are shown in Figs. 1 and 2. Source III is a simulated source consisting of two point sources convolved with a circularly symmetric Gaussian distribution, with 1% noise added. The output map is shown in Fig. 3.

The computer time needed for each source on the Caltech computer "PHOBOS" was 10-15 minutes, which compares favorably with the time for conventional methods. This is possible because the new method can give good maps with a coarser grid than other methods. This is shown by Fig. 4, which displays the map of Source I derived in the usual way with a fine grid, and the effect on this map if a 32×32 grid is used for the final inversion and "CLEANing" ("CLEAN" is a Caltech program).

The method described here has been shown to be a practical alternative to existing methods. It can construct a map without any user intervention. The resulting map is free of the biases introduced by interpolation before inversion and by the CLEAN program.

Acknowledgment

I am indebted to Dr. T. J. Pearson of the Caltech Astronomy Department for help in providing data and access to the Campus/JPL Correlation Facility located in Robinson at Caltech.

References

1. Readhead, A. C. S., and Wilkinson, P. N., "The mapping of compact radio sources from VLBI data," *Astrophysical Journal*, 223, pp. 25-36 (1978).
2. Cotton, W. D., "A method of mapping compact structure in radio sources using VLBI observations," *Astronomical Journal*, 84, pp. 1122-1128 (1979).
3. Readhead, A. C. S., Walker, R. C., Pearson, T. J., and Cohen, M. H., "Mapping radio sources with uncalibrated visibility data," *Nature*, 285, pp. 137-140 (1980).
4. Cornwell, T. J., and Wilkinson, P. N., "A new method for making maps with unstable radio interferometers," *Mon. Not. R. Astro. Soc.*, 196, pp. 1067-1086 (1981).

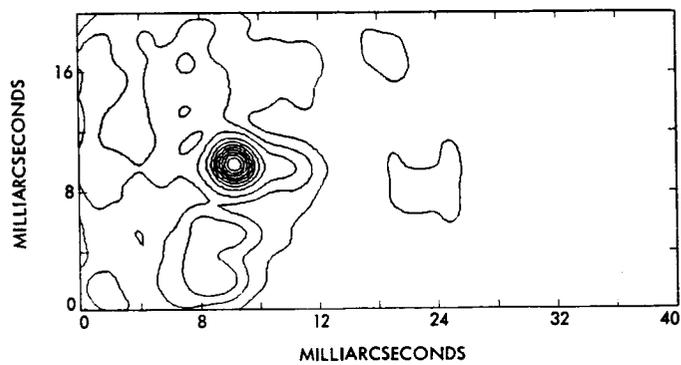


Fig. 1. Source 1807+698

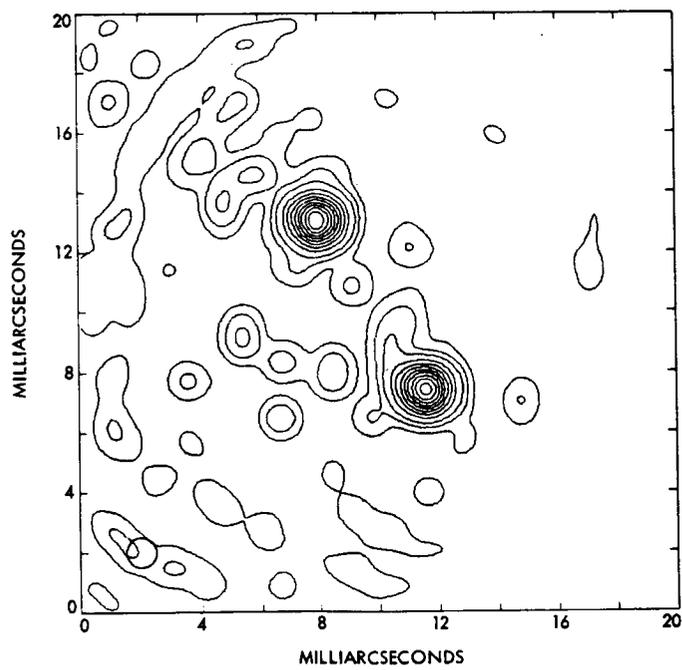


Fig. 2. Source 2021+614

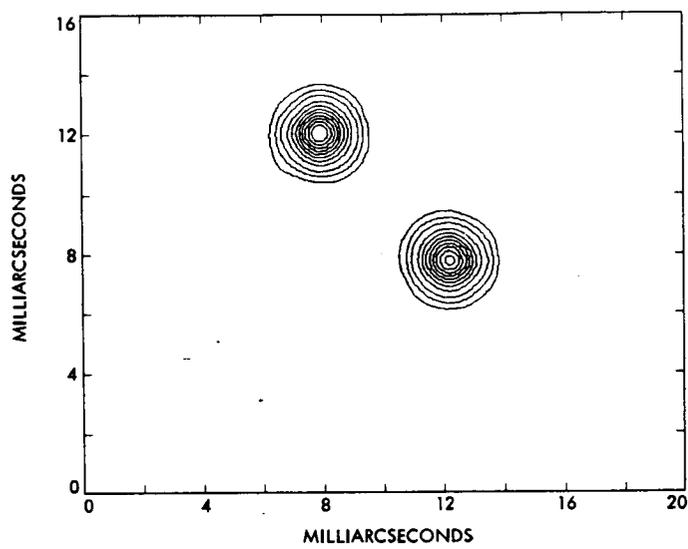


Fig. 3. Simulated source

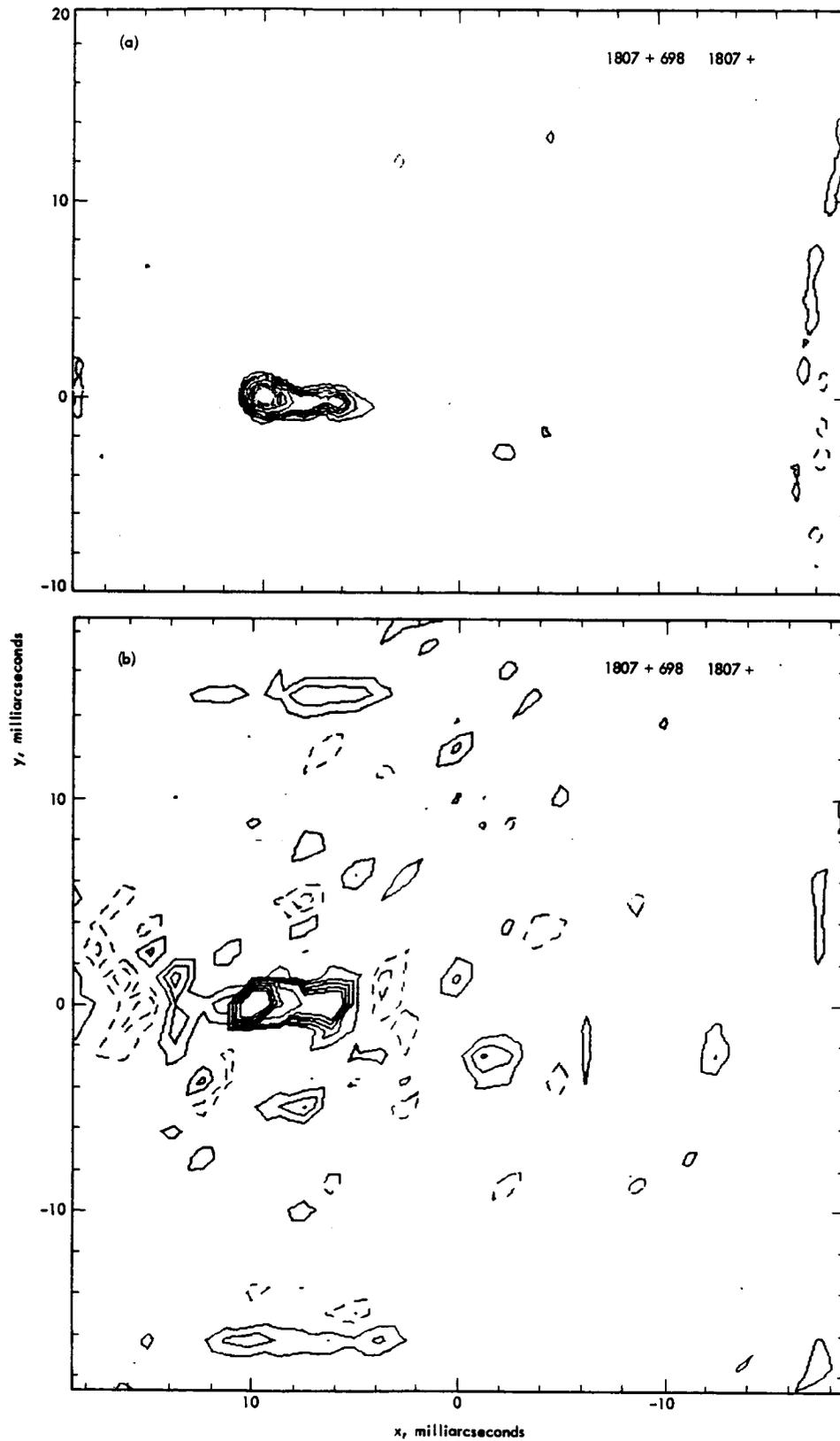


Fig. 4. Output of CLEAN: (a) 128×128 resolution; (b) 32×32 resolution

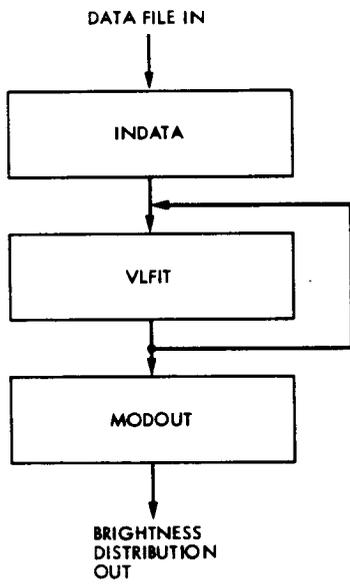


Fig. 5. Relation of the programs

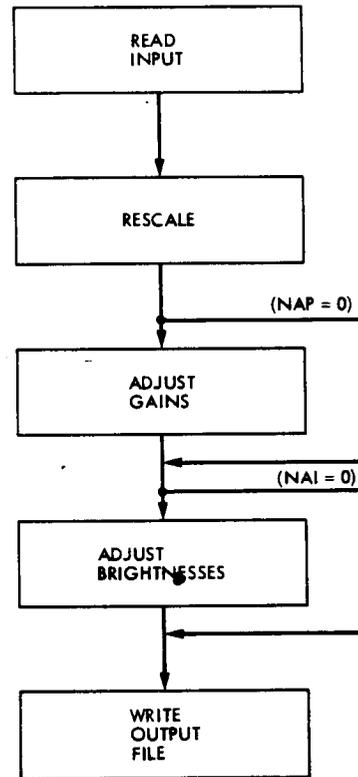


Fig. 6. The program VLFIT

Appendix A

Derivation of the Cost Function

In determining the brightness distribution of a radio source, preliminary reduction of VLBI data yields corrupted values of the visibility function

$$V(u, v) = \iint I(x, y) \exp [2\pi i(ux + vy)] dx dy \quad (\text{A1})$$

[Refs. A-1 and A-2]. Here $I(x, y)$ is the brightness distribution, expressed as a function of rectangular coordinates x, y in the tangent plane to the celestial sphere at the source location. The visibility function $V(u, v)$ is its two-dimensional Fourier transform.

The core of the problem of finding the brightness distribution is the inversion of this transform. This is complicated by the fact that only values of $V(u, v)$ on a restricted, irregularly spaced set of points (u, v) are contained in the data.

Suppose there are NS antennas. Then the quantities u_p, v_p are associated with the p^{th} antenna, $1 \leq p \leq NS$, as follows. Let X_p be the vector from the center of the earth to the p^{th} antenna, and λ be the wavelength of the radiation received. Let K be the factor by which values of x, y must be multiplied to convert to radians (typically, x and y are measured in milliarcseconds, and $K = \pi \cdot 10^{-6}/648$). Then u_p, v_p are the components of KX_p/λ parallel to the x and y axes.

The only values of $V(u, v)$ which enter into the data are

$$V_{pq} = V(u_p - u_q, v_p - v_q), \quad 1 \leq p, q \leq NS.$$

These quantities are functions of time, since u_p, v_p change due to the rotation of the earth. The data are averaged over time intervals sufficiently small that there is no significant variation of u_p, v_p over an interval. This reduces the set of visibilities to a finite set of values $V_{pq}^{(n)}$, where n is the index of the time interval. For a given value of n , some pairs p, q may not occur, because the source was not visible from certain antennas at that time, or because some data were lost.

To determine $I(x, y)$ numerically, it is approximated by a series of delta functions on a rectangular grid in a restricted region of the plane,

$$I(x, y) = \sum_{m=1}^{MF} I_m \delta(x - x_m) \delta(y - y_m).$$

Then equation (A1) is reduced to

$$V_{pq}^{(n)} = \sum_{m=1}^{MF} C_{pq,m}^{(n)} I_m, \quad (\text{A2})$$

where

$$C_{pq,m}^{(n)} = \exp [2\pi i \{ (u_p^{(n)} - u_q^{(n)}) x_m + (v_p^{(n)} - v_q^{(n)}) y_m \}].$$

Unfortunately, there are additional complications in the problem. Each antenna has a gain, M_p , and a phase offset, ϕ_p , which are unknown functions of time. The effects of noise must also be taken into consideration. If $A_p^{(n)} = M_p \exp(i\phi_p)$, then the quantities actually given by the data are

$$E_{pq}^{(n)} = A_p^{(n)} \overline{A_q^{(n)}} V_{pq}^{(n)} + n_{pq}^{(n)} \quad (\text{A3})$$

where the bar denotes a complex conjugate, and $n_{pq}^{(n)}$ is the contribution of the noise [Ref. 3].

Our problem is to determine the values I_m , given the quantities $E_{pq}^{(n)}$. If we start with assumed values $\hat{A}_p^{(n)}, \hat{I}_m$, we can construct

$$\hat{V}_{pq}^{(n)} = \sum_{m=1}^{MF} C_{pq,m}^{(n)} \hat{I}_m,$$

and

$$F_{pq}^{(n)} = \hat{A}_p^{(n)} \overline{\hat{A}_q^{(n)}} \hat{V}_{pq}^{(n)}. \quad (\text{A4})$$

The noise terms $n_{pq}^{(n)}$ in (A3) are assumed to be unknown, so a criterion for the goodness of the assumed values is how close the $F_{pq}^{(n)}$ are to the $E_{pq}^{(n)}$. The averaging process which furnishes the data $E_{pq}^{(n)}$ also provides estimates, $\sigma_{pq}^{(n)}$, of the standard deviation of the noise. Thus we are led to consider the cost function

$$Q = \sum_n \sum_{p,q} \left| F_{pq}^{(n)} - E_{pq}^{(n)} \right|^2 / \sigma_{pq}^{(n)2}. \quad (\text{A5})$$

Minimization of Q gives the least-squares fit of the $F_{pq}^{(n)}$ to the $E_{pq}^{(n)}$.

References

- A-1. Pachoczyk, A. G., *Radio Astrophysics*, Freeman, San Francisco (1970).
- A-2. Fomalont, E. B., and Wright, M. C. H., in *Galactic and Extragalactic Radio Astronomy* (eds. Verschunt, G. L. and Kellermann, K. I.), Springer, New York (1974).

Appendix B

Iterative Minimization of the Cost Function

As stated in Section II, this minimization is accomplished by iterating two steps.

I. Variation of \hat{I}_m

Let $z_m = \hat{I}_m$. Then, in its dependence on the z_m , Q is a positive definite quadratic form

$$Q = \sum_{m,k=1}^{MF} B_{mk} z_m z_k - 2 \sum_{m=1}^{MF} D_m z_m + R, \quad (\text{B1})$$

where

$$B_{mk} = \text{Real} \left\{ \sum_n \sum_{p,q} \left| \hat{A}_p^{(n)} \hat{A}_q^{(n)} \right|^2 \frac{C_{pq,m}^{(n)} \overline{C_{pq,k}^{(n)}}}{\sigma_{pq}^{(n)2}} \right\},$$

$$D_m = \text{Real} \left\{ \sum_n \sum_{p,q} A_p^{(n)} \overline{A_q^{(n)}} \frac{C_{pq,m}^{(n)} \overline{E_{pq}^{(n)}}}{\sigma_{pq}^{(n)2}} \right\},$$

and

$$R = \sum_n \sum_{p,q} \left| E_{pq}^{(n)} \right|^2 / \sigma_{pq}^{(n)2}.$$

This quadratic form is minimized by the values of the z_m at which all the partial derivatives with respect to the z_m are zero:

$$\sum_{k=1}^{MF} B_{mk} z_k = D_m, \quad 1 \leq m \leq MF \quad (\text{B2})$$

Direct solution of this system is not practical because of the large number of unknowns. An iterative method which converges to the solution is the *Gauss-Seidel* method [Ref. B-1], which consists of the following: Start with any assumed values of the unknowns. For $m = 1$ to MF , make the replacement

$$z_m \leftarrow \left(D_m - \sum_{k \neq m} B_{mk} z_k \right) / B_{mm}. \quad (\text{B3})$$

When applied repeatedly, this procedure is easily shown to converge to the minimum of Q (for fixed $\hat{A}_p^{(n)}$). The step (B3) gives the minimum of Q when only the one unknown z_m is allowed to vary.

II. Variation of the $A_p(n)$

Here each of the subsums of (A5),

$$Q_n = \sum_{p,q} \left| F_{pq}^{(n)} - E_{pq}^{(n)} \right|^2 / \sigma_{pq}^{(n)2},$$

involves a separate set of variables, so they can be minimized separately. Let $a_p = \hat{A}_p^{(n)}$. Then

$$Q_n = \sum_{p,q} \left(G_{pq} |a_p a_q|^2 - 2H_{pq} a_p \overline{a_q} \right) + R_n, \quad (\text{B4})$$

where

$$G_{pq} = \left| \hat{V}_{pq}^{(n)} \right|^2 / \sigma_{pq}^{(n)2},$$

$$H_{pq} = \hat{V}_{pq}^{(n)} \overline{E_{pq}^{(n)}} / \sigma_{pq}^{(n)2},$$

and

$$R_n = \sum_{p,q} \left| E_{pq}^{(n)} \right|^2 / \sigma_{pq}^{(n)2}.$$

This function is minimized by the gradient method [Ref. B-1]. Let

$$b_p = - \frac{\partial Q_n}{\partial a_p} = 2 \sum_q \left(G_{pq} |a_q|^2 a_p + H_{qp} a_q \right).$$

For $\nu > 0$, the quantities $a_p + \nu b_p$ are directed away from a_p along the negative gradient of Q_n (in the $2 \cdot NS$ -dimensional space of the real and imaginary parts of the a_p 's). The iterative step here is to replace a_p by $a_p + \nu b_p$ in the expression (B4) for Q_n , then choose ν to minimize the resulting fourth degree

polynomial in v . Using the minimizing value, $a_p + vb_p$ gives the new value of a_p .

In Step (2), as in (1), convergence is guaranteed since each step decreases Q . More generally, the steps of (1) and (2) can be intermingled in any systematic iterative procedure which

varies each unknown infinitely often, and Q will approach the minimum. It is advantageous, however, to carry out Steps (1) and (2) separately to the point where the minimums of the separated problems are approached, because this reduces the amount of time spent in computing the coefficients in the formulas (B1) and (B4).

Reference

- B-1. Burden, R. L., Faires, J. D., and Reynolds, A. C., *Numerical Analysis*, Prindle, Weber & Schmidt, Boston (1978).

Appendix C

Description of the Computer Programs

The method described here was implemented on the VAX computer PHOBOS in the Caltech Astronomy department. A series of three programs is used (see Fig. 5):

- (1) The program INDATA reads the data from an existing data file in the MERGE format currently used at Caltech. This data is transferred into new files, with some conversion, and some of the quantities to be used in the minimization process are pre-computed.
- (2) The program VLFIT carries out the minimization procedure of Section II.
- (3) The program MODOUT reads the file output by VLFIT and builds a new file MOD.DAT which can be used by the Caltech program MODPLOT to draw a contour map of the source.

I. Program INDATA

This program reads an input file prepared by other existing programs. The first section reads a collection of header records, saving some of the information and converting some to a more convenient form. This information includes the astronomical location of the source (right ascension and declination), the starting time of the data, and names and locations of all the antennas. The time origin is chosen to be the start of the day on which the data begins, in Greenwich mean time. The right ascension is converted to $PLO\emptyset$, the longitude of the source at time \emptyset . The vector from the earth's center to the p^{th} antenna, $(STX(p), STY(p), STZ(p))$, is rescaled so that its components in the xy -plane at the source are the quantities $2\pi u_p, 2\pi v_p$ of Appendix A.

Next the data section of the file is read. The data points are stored sequentially in an array of length IP . The antenna pair (I, J) is denoted by one index $L = L(I, J)$ which is related to I and J by the arrays IIB, JIB set up in the program. We build the arrays

$$T(N), \quad N=1, \dots, NT$$

and

$$E(K), SGM(K), LP(K), KT(K), \quad K=1, \dots, IP,$$

where $T(N)$ is the time of the n^{th} time interval in minutes, and

$$E(K) = E_{pq}^{(N)},$$

$$SGM(K) = \sigma_{pq}^{(N)^2},$$

$$LP(K) = L(p, q),$$

with

$$KT(K) = N, \text{ in the } n^{\text{th}} \text{ time interval.}$$

Data which contain no useful information are eliminated.

Next the program reads two lines of input data which specify parameters of the map. These are:

- (1) NX, NY : Number of grid points in the x and y directions.
- (2) XL, YL : Half-width of the map in x and y directions, in milliarcseconds.

The total number of real unknowns is now known to be $2NS \cdot NT + MF$ (where $MF = NX \cdot NY$). The number of real conditions in the data is $2 \cdot IP$. The ratio $2IP / (2NS \cdot NT + MF)$, the redundancy, is evaluated here. If this value is less than 1, there are not enough conditions to determine the unknowns. If this happens, we may still get a good map, but there are other solutions to the minimum problem.

The next section of the program computes the quantities $2\pi u_p, 2\pi v_p$ for each time interval, and forms the auxiliary complex arrays $CL\emptyset(p, N), CLX(p, N), CLY(p, N)$. These are used in VLFIT to generate $C_{pq, m}^{(n)}$ as follows: The x and y values at the grid points are

$$x_M = x_o + (I - 1)DX, \quad 1 \leq I \leq NX,$$

$$y_M = y_o + (J - 1)DY, \quad 1 \leq J \leq NY,$$

where $M = J + (I - 1)NY$. If we put

$$CL(p, M, N) = CL\emptyset(p, N) \cdot CLX(p, N)^{I-1} \cdot CLY(p, N)^{J-1},$$

then

$$C_{pq, M}^{(N)} = CL(p, M, N) \cdot \overline{CL(q, M, N)}.$$

Finally, the brightness array $AI(M)$ and the complex gain array $AP(I, N)$ are initialized with values 1, and all the arrays needed by VLFIT are written in files.

II. Program VLFIT

This program reads the files prepared by INDATA, and two other lines of input: (i) NAP and NAI, and (ii) DSC (see Fig. 6). The gains are not adjusted if NAP = 0. The brightnesses are only rescaled if NAI = 0. DSC is a parameter used in image enhancement. For DSC > 0, the peaks of the brightness distribution are sharpened twice during the iterative solution for the brightnesses $AI(M)$.

First the brightnesses are rescaled so that for the given gains and shape of the brightness distribution, the fit to the data is as good as possible.

Next (if NAP≠0) step (2) of Appendix B is applied, with twelve iterations in each time interval.

Next (if NAI≠0) step (1) is applied. The values of $AI(M)$, $M=1$ to MF , are adjusted fifty times. After the 6th time and the 26th time, image enhancement is applied, by replacing each $AI(M)$ by

$$AI(M) - DSC \cdot (\max(AI) - AI(M))$$

and then replacing any negative values by zero.

Finally, an output file of values of gains and brightnesses is written, in the same format as the input file. This allows the application of VLFIT any number of times, using each output file as input the next time.

III. Program MODOUT

This program converts the output of VLFIT into a form useable by the program MODPLOT. The brightness distribution is characterized by a series of components, in this case delta-functions at the grid points, whose position is given in polar coordinates. These are listed in order of decreasing strength, terminating after at most MAXC components, where MAXC is a number read as input by the program.

Referees

The following people have refereed articles for *The Telecommunications and Data Acquisition Progress Report*. By attesting to the technical and archival value of the articles, they have helped to maintain the excellence of this publication during the past year.

| | | | |
|-----------------|------------------|--------------------|-------------------|
| D. A. Bathker | C. E. Hildebrand | J. R. Lesh | A. M. Ruskin |
| J. J. Bautista | R. L. Horttor | R. Levy | R. Sfeir |
| D. W. Brown | D. L. Johnson | K. M. Liewer | H. M. Shao |
| R. W. Burt | J. Katz | G. Lorden | R. Stevens |
| A. G. Cha | M. J. Klein | R. J. McEliece | L. Swanson |
| R. B. Crow | C. E. Kohlhase | S. N. Mohan | R. N. Treuhaft |
| L. J. Deutsch | R. Kumar | B. Moultrie | C. Wang |
| J. L. Faselow | C. R. Lahmeyer | H. P. Phillips | R. A. Winkelstein |
| E. C. Gatz | D. B. Lane | F. Pollara-Bozzola | S. S. Wissler |
| R. M. Goldstein | F. L. Lansing | G. H. Purcell, Jr. | M. Wyatt |
| C. A. Greenhall | G. E. Lanyi | M. P. Quirk | C. F. Yoder |
| R. W. Hartop | J. W. Layland | D. L. Robinson | |